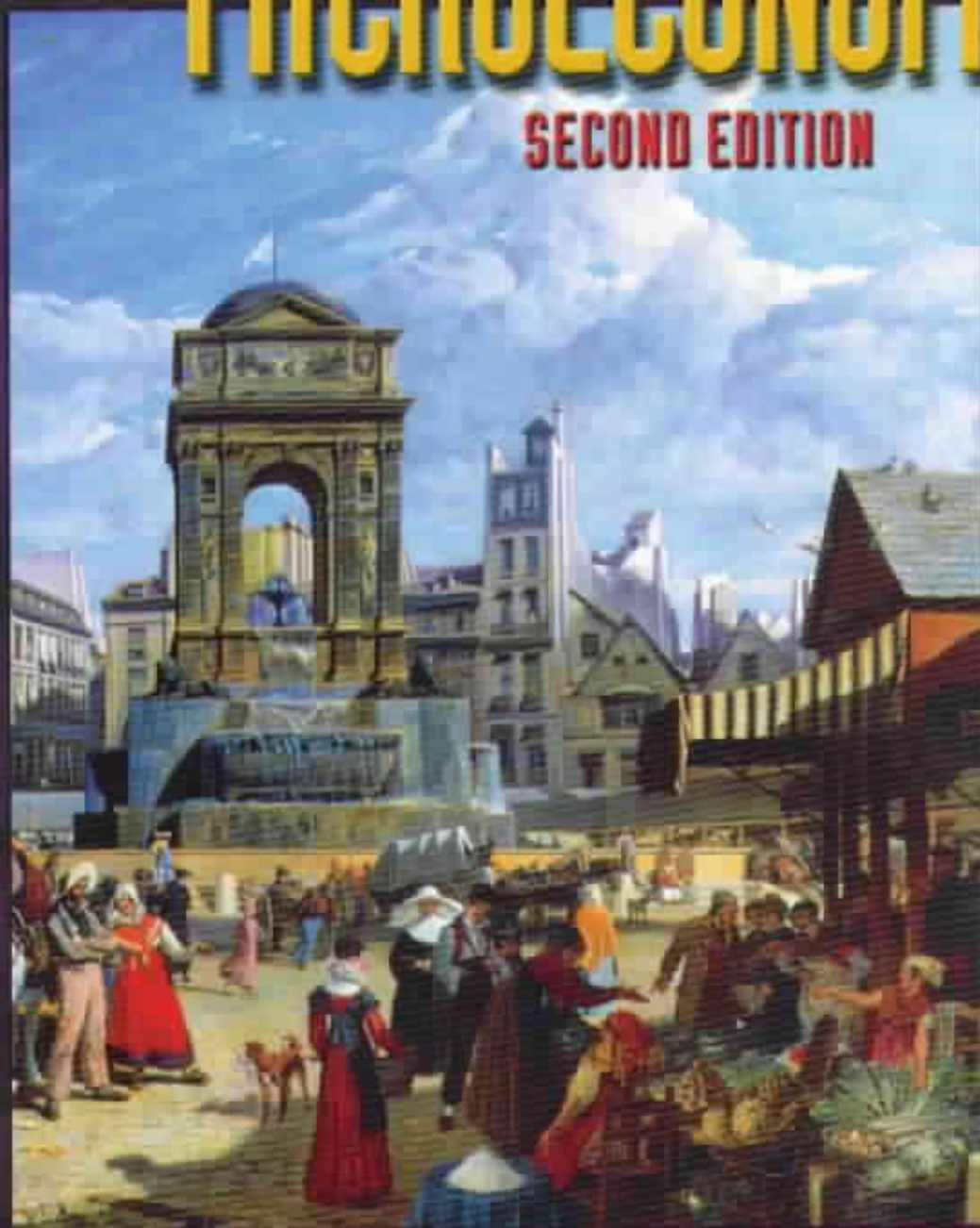


PRINCIPLES OF MICROECONOMICS

SECOND EDITION



N. GREGORY MANKIW

1



TEN PRINCIPLES OF ECONOMICS

The word *economy* comes from the Greek word for “one who manages a household.” At first, this origin might seem peculiar. But, in fact, households and economies have much in common.

A household faces many decisions. It must decide which members of the household do which tasks and what each member gets in return: Who cooks dinner? Who does the laundry? Who gets the extra dessert at dinner? Who gets to choose what TV show to watch? In short, the household must allocate its scarce resources among its various members, taking into account each member’s abilities, efforts, and desires.

Like a household, a society faces many decisions. A society must decide what jobs will be done and who will do them. It needs some people to grow food, other people to make clothing, and still others to design computer software. Once society has allocated people (as well as land, buildings, and machines) to various jobs,

IN THIS CHAPTER YOU WILL . . .

Learn that economics is about the allocation of scarce resources

Examine some of the tradeoffs that people face

Learn the meaning of opportunity cost

See how to use marginal reasoning when making decisions

Discuss how incentives affect people’s behavior

Consider why trade among people or nations can be good for everyone

Discuss why markets are a good, but not perfect, way to allocate resources

Learn what determines some trends in the overall economy

scarcity

the limited nature of society's resources

economics

the study of how society manages its scarce resources



it must also allocate the output of goods and services that they produce. It must decide who will eat caviar and who will eat potatoes. It must decide who will drive a Porsche and who will take the bus.

The management of society's resources is important because resources are scarce. **Scarcity** means that society has limited resources and therefore cannot produce all the goods and services people wish to have. Just as a household cannot give every member everything he or she wants, a society cannot give every individual the highest standard of living to which he or she might aspire.

Economics is the study of how society manages its scarce resources. In most societies, resources are allocated not by a single central planner but through the combined actions of millions of households and firms. Economists therefore study how people make decisions: how much they work, what they buy, how much they save, and how they invest their savings. Economists also study how people interact with one another. For instance, they examine how the multitude of buyers and sellers of a good together determine the price at which the good is sold and the quantity that is sold. Finally, economists analyze forces and trends that affect the economy as a whole, including the growth in average income, the fraction of the population that cannot find work, and the rate at which prices are rising.

Although the study of economics has many facets, the field is unified by several central ideas. In the rest of this chapter, we look at *Ten Principles of Economics*. These principles recur throughout this book and are introduced here to give you an overview of what economics is all about. You can think of this chapter as a “preview of coming attractions.”

HOW PEOPLE MAKE DECISIONS

There is no mystery to what an “economy” is. Whether we are talking about the economy of Los Angeles, of the United States, or of the whole world, an economy is just a group of people interacting with one another as they go about their lives. Because the behavior of an economy reflects the behavior of the individuals who make up the economy, we start our study of economics with four principles of individual decisionmaking.

PRINCIPLE #1: PEOPLE FACE TRADEOFFS

The first lesson about making decisions is summarized in the adage: “There is no such thing as a free lunch.” To get one thing that we like, we usually have to give up another thing that we like. Making decisions requires trading off one goal against another.

Consider a student who must decide how to allocate her most valuable resource—her time. She can spend all of her time studying economics; she can spend all of her time studying psychology; or she can divide her time between the two fields. For every hour she studies one subject, she gives up an hour she could have used studying the other. And for every hour she spends studying, she gives up an hour that she could have spent napping, bike riding, watching TV, or working at her part-time job for some extra spending money.

Or consider parents deciding how to spend their family income. They can buy food, clothing, or a family vacation. Or they can save some of the family income for retirement or the children's college education. When they choose to spend an extra dollar on one of these goods, they have one less dollar to spend on some other good.

When people are grouped into societies, they face different kinds of tradeoffs. The classic tradeoff is between "guns and butter." The more we spend on national defense to protect our shores from foreign aggressors (guns), the less we can spend on consumer goods to raise our standard of living at home (butter). Also important in modern society is the tradeoff between a clean environment and a high level of income. Laws that require firms to reduce pollution raise the cost of producing goods and services. Because of the higher costs, these firms end up earning smaller profits, paying lower wages, charging higher prices, or some combination of these three. Thus, while pollution regulations give us the benefit of a cleaner environment and the improved health that comes with it, they have the cost of reducing the incomes of the firms' owners, workers, and customers.

Another tradeoff society faces is between efficiency and equity. **Efficiency** means that society is getting the most it can from its scarce resources. **Equity** means that the benefits of those resources are distributed fairly among society's members. In other words, efficiency refers to the size of the economic pie, and equity refers to how the pie is divided. Often, when government policies are being designed, these two goals conflict.

Consider, for instance, policies aimed at achieving a more equal distribution of economic well-being. Some of these policies, such as the welfare system or unemployment insurance, try to help those members of society who are most in need. Others, such as the individual income tax, ask the financially successful to contribute more than others to support the government. Although these policies have the benefit of achieving greater equity, they have a cost in terms of reduced efficiency. When the government redistributes income from the rich to the poor, it reduces the reward for working hard; as a result, people work less and produce fewer goods and services. In other words, when the government tries to cut the economic pie into more equal slices, the pie gets smaller.

Recognizing that people face tradeoffs does not by itself tell us what decisions they will or should make. A student should not abandon the study of psychology just because doing so would increase the time available for the study of economics. Society should not stop protecting the environment just because environmental regulations reduce our material standard of living. The poor should not be ignored just because helping them distorts work incentives. Nonetheless, acknowledging life's tradeoffs is important because people are likely to make good decisions only if they understand the options that they have available.

efficiency

the property of society getting the most it can from its scarce resources

equity

the property of distributing economic prosperity fairly among the members of society

PRINCIPLE #2: THE COST OF SOMETHING IS WHAT YOU GIVE UP TO GET IT

Because people face tradeoffs, making decisions requires comparing the costs and benefits of alternative courses of action. In many cases, however, the cost of some action is not as obvious as it might first appear.

Consider, for example, the decision whether to go to college. The benefit is intellectual enrichment and a lifetime of better job opportunities. But what is the cost? To answer this question, you might be tempted to add up the money you

spend on tuition, books, room, and board. Yet this total does not truly represent what you give up to spend a year in college.

The first problem with this answer is that it includes some things that are not really costs of going to college. Even if you quit school, you would need a place to sleep and food to eat. Room and board are costs of going to college only to the extent that they are more expensive at college than elsewhere. Indeed, the cost of room and board at your school might be less than the rent and food expenses that you would pay living on your own. In this case, the savings on room and board are a benefit of going to college.

The second problem with this calculation of costs is that it ignores the largest cost of going to college—your time. When you spend a year listening to lectures, reading textbooks, and writing papers, you cannot spend that time working at a job. For most students, the wages given up to attend school are the largest single cost of their education.

opportunity cost

whatever must be given up to obtain some item

The **opportunity cost** of an item is what you give up to get that item. When making any decision, such as whether to attend college, decisionmakers should be aware of the opportunity costs that accompany each possible action. In fact, they usually are. College-age athletes who can earn millions if they drop out of school and play professional sports are well aware that their opportunity cost of college is very high. It is not surprising that they often decide that the benefit is not worth the cost.

PRINCIPLE #3: RATIONAL PEOPLE THINK AT THE MARGIN

Decisions in life are rarely black and white but usually involve shades of gray. When it's time for dinner, the decision you face is not between fasting or eating like a pig, but whether to take that extra spoonful of mashed potatoes. When exams roll around, your decision is not between blowing them off or studying 24 hours a day, but whether to spend an extra hour reviewing your notes instead of watching TV. Economists use the term **marginal changes** to describe small incremental adjustments to an existing plan of action. Keep in mind that "margin" means "edge," so marginal changes are adjustments around the edges of what you are doing.

marginal changes

small incremental adjustments to a plan of action

In many situations, people make the best decisions by thinking at the margin. Suppose, for instance, that you asked a friend for advice about how many years to stay in school. If he were to compare for you the lifestyle of a person with a Ph.D. to that of a grade school dropout, you might complain that this comparison is not helpful for your decision. You have some education already and most likely are deciding whether to spend an extra year or two in school. To make this decision, you need to know the additional benefits that an extra year in school would offer (higher wages throughout life and the sheer joy of learning) and the additional costs that you would incur (tuition and the forgone wages while you're in school). By comparing these *marginal benefits* and *marginal costs*, you can evaluate whether the extra year is worthwhile.

As another example, consider an airline deciding how much to charge passengers who fly standby. Suppose that flying a 200-seat plane across the country costs the airline \$100,000. In this case, the average cost of each seat is \$100,000/200, which is \$500. One might be tempted to conclude that the airline should never sell a ticket for less than \$500. In fact, however, the airline can raise its profits by

thinking at the margin. Imagine that a plane is about to take off with ten empty seats, and a standby passenger is waiting at the gate willing to pay \$300 for a seat. Should the airline sell it to him? Of course it should. If the plane has empty seats, the cost of adding one more passenger is minuscule. Although the *average* cost of flying a passenger is \$500, the *marginal* cost is merely the cost of the bag of peanuts and can of soda that the extra passenger will consume. As long as the standby passenger pays more than the marginal cost, selling him a ticket is profitable.

As these examples show, individuals and firms can make better decisions by thinking at the margin. A rational decisionmaker takes an action if and only if the marginal benefit of the action exceeds the marginal cost.

PRINCIPLE #4: PEOPLE RESPOND TO INCENTIVES

Because people make decisions by comparing costs and benefits, their behavior may change when the costs or benefits change. That is, people respond to incentives. When the price of an apple rises, for instance, people decide to eat more pears and fewer apples, because the cost of buying an apple is higher. At the same time, apple orchards decide to hire more workers and harvest more apples, because the benefit of selling an apple is also higher. As we will see, the effect of price on the behavior of buyers and sellers in a market—in this case, the market for apples—is crucial for understanding how the economy works.

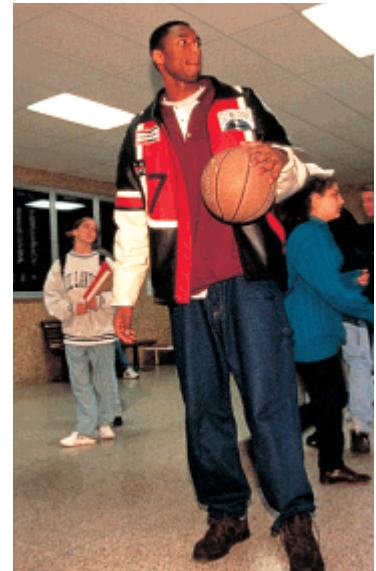
Public policymakers should never forget about incentives, for many policies change the costs or benefits that people face and, therefore, alter behavior. A tax on gasoline, for instance, encourages people to drive smaller, more fuel-efficient cars. It also encourages people to take public transportation rather than drive and to live closer to where they work. If the tax were large enough, people would start driving electric cars.

When policymakers fail to consider how their policies affect incentives, they can end up with results that they did not intend. For example, consider public policy regarding auto safety. Today all cars have seat belts, but that was not true 40 years ago. In the late 1960s, Ralph Nader's book *Unsafe at Any Speed* generated much public concern over auto safety. Congress responded with laws requiring car companies to make various safety features, including seat belts, standard equipment on all new cars.

How does a seat belt law affect auto safety? The direct effect is obvious. With seat belts in all cars, more people wear seat belts, and the probability of surviving a major auto accident rises. In this sense, seat belts save lives.

But that's not the end of the story. To fully understand the effects of this law, we must recognize that people change their behavior in response to the incentives they face. The relevant behavior here is the speed and care with which drivers operate their cars. Driving slowly and carefully is costly because it uses the driver's time and energy. When deciding how safely to drive, rational people compare the marginal benefit from safer driving to the marginal cost. They drive more slowly and carefully when the benefit of increased safety is high. This explains why people drive more slowly and carefully when roads are icy than when roads are clear.

Now consider how a seat belt law alters the cost-benefit calculation of a rational driver. Seat belts make accidents less costly for a driver because they reduce the probability of injury or death. Thus, a seat belt law reduces the benefits to slow and careful driving. People respond to seat belts as they would to an improvement



BASKETBALL STAR KOBE BRYANT UNDERSTANDS OPPORTUNITY COST AND INCENTIVES. DESPITE GOOD HIGH SCHOOL GRADES AND SAT SCORES, HE DECIDED TO SKIP COLLEGE AND GO STRAIGHT TO THE NBA, WHERE HE EARNED ABOUT \$10 MILLION OVER FOUR YEARS.

in road conditions—by faster and less careful driving. The end result of a seat belt law, therefore, is a larger number of accidents.

How does the law affect the number of deaths from driving? Drivers who wear their seat belts are more likely to survive any given accident, but they are also more likely to find themselves in an accident. The net effect is ambiguous. Moreover, the reduction in safe driving has an adverse impact on pedestrians (and on drivers who do not wear their seat belts). They are put in jeopardy by the law because they are more likely to find themselves in an accident but are not protected by a seat belt. Thus, a seat belt law tends to increase the number of pedestrian deaths.

At first, this discussion of incentives and seat belts might seem like idle speculation. Yet, in a 1975 study, economist Sam Peltzman showed that the auto-safety laws have, in fact, had many of these effects. According to Peltzman's evidence, these laws produce both fewer deaths per accident and more accidents. The net result is little change in the number of driver deaths and an increase in the number of pedestrian deaths.

Peltzman's analysis of auto safety is an example of the general principle that people respond to incentives. Many incentives that economists study are more straightforward than those of the auto-safety laws. No one is surprised that people drive smaller cars in Europe, where gasoline taxes are high, than in the United States, where gasoline taxes are low. Yet, as the seat belt example shows, policies can have effects that are not obvious in advance. When analyzing any policy, we must consider not only the direct effects but also the indirect effects that work through incentives. If the policy changes incentives, it will cause people to alter their behavior.

QUICK QUIZ: List and briefly explain the four principles of individual decisionmaking.

HOW PEOPLE INTERACT

The first four principles discussed how individuals make decisions. As we go about our lives, many of our decisions affect not only ourselves but other people as well. The next three principles concern how people interact with one another.

PRINCIPLE #5: TRADE CAN MAKE EVERYONE BETTER OFF

You have probably heard on the news that the Japanese are our competitors in the world economy. In some ways, this is true, for American and Japanese firms do produce many of the same goods. Ford and Toyota compete for the same customers in the market for automobiles. Compaq and Toshiba compete for the same customers in the market for personal computers.

Yet it is easy to be misled when thinking about competition among countries. Trade between the United States and Japan is not like a sports contest, where one

side wins and the other side loses. In fact, the opposite is true: Trade between two countries can make each country better off.

To see why, consider how trade affects your family. When a member of your family looks for a job, he or she competes against members of other families who are looking for jobs. Families also compete against one another when they go shopping, because each family wants to buy the best goods at the lowest prices. So, in a sense, each family in the economy is competing with all other families.

Despite this competition, your family would not be better off isolating itself from all other families. If it did, your family would need to grow its own food, make its own clothes, and build its own home. Clearly, your family gains much from its ability to trade with others. Trade allows each person to specialize in the activities he or she does best, whether it is farming, sewing, or home building. By trading with others, people can buy a greater variety of goods and services at lower cost.

Countries as well as families benefit from the ability to trade with one another. Trade allows countries to specialize in what they do best and to enjoy a greater variety of goods and services. The Japanese, as well as the French and the Egyptians and the Brazilians, are as much our partners in the world economy as they are our competitors.

PRINCIPLE #6: MARKETS ARE USUALLY A GOOD WAY TO ORGANIZE ECONOMIC ACTIVITY

The collapse of communism in the Soviet Union and Eastern Europe may be the most important change in the world during the past half century. Communist countries worked on the premise that central planners in the government were in the best position to guide economic activity. These planners decided what goods and services were produced, how much was produced, and who produced and consumed these goods and services. The theory behind central planning was that only the government could organize economic activity in a way that promoted economic well-being for the country as a whole.

Today, most countries that once had centrally planned economies have abandoned this system and are trying to develop market economies. In a **market economy**, the decisions of a central planner are replaced by the decisions of millions of firms and households. Firms decide whom to hire and what to make. Households decide which firms to work for and what to buy with their incomes. These firms and households interact in the marketplace, where prices and self-interest guide their decisions.

At first glance, the success of market economies is puzzling. After all, in a market economy, no one is looking out for the economic well-being of society as a whole. Free markets contain many buyers and sellers of numerous goods and services, and all of them are interested primarily in their own well-being. Yet, despite decentralized decisionmaking and self-interested decisionmakers, market economies have proven remarkably successful in organizing economic activity in a way that promotes overall economic well-being.

In his 1776 book *An Inquiry into the Nature and Causes of the Wealth of Nations*, economist Adam Smith made the most famous observation in all of economics: Households and firms interacting in markets act as if they are guided by an “invisible hand” that leads them to desirable market outcomes. One of our goals in



“For \$5 a week you can watch baseball without being nagged to cut the grass!”

market economy

an economy that allocates resources through the decentralized decisions of many firms and households as they interact in markets for goods and services

FYI

Adam Smith
and the
Invisible Hand



hand of government guiding their actions. This political philosophy provides the intellectual basis for the market economy, and for free society more generally.

Why do decentralized market economies work so well? Is it because people can be counted on to treat one another with love and kindness? Not at all. Here is Adam Smith's description of how people interact in a market economy:

Man has almost constant occasion for the help of his brethren, and it is vain for him to expect it from their benevolence only. He will be more likely to prevail if he can interest their self-love in his favor, and show them that it is for their own advantage to do for him what he requires of them. . . . It is not from the benevolence of

It may be only a coincidence that Adam Smith's great book, *An Inquiry into the Nature and Causes of the Wealth of Nations*, was published in 1776, the exact year American revolutionaries signed the Declaration of Independence. But the two documents do share a point of view that was prevalent at the time—that individuals are usually best left to their own devices, without the heavy

the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest. . . .

Every individual . . . neither intends to promote the public interest, nor knows how much he is promoting it. . . . He intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it.



ADAM SMITH

Smith is saying that participants in the economy are motivated by self-interest and that the “invisible hand” of the marketplace guides this self-interest into promoting general economic well-being.

Many of Smith's insights remain at the center of modern economics. Our analysis in the coming chapters will allow us to express Smith's conclusions more precisely and to analyze fully the strengths and weaknesses of the market's invisible hand.

this book is to understand how this invisible hand works its magic. As you study economics, you will learn that prices are the instrument with which the invisible hand directs economic activity. Prices reflect both the value of a good to society and the cost to society of making the good. Because households and firms look at prices when deciding what to buy and sell, they unknowingly take into account the social benefits and costs of their actions. As a result, prices guide these individual decisionmakers to reach outcomes that, in many cases, maximize the welfare of society as a whole.

There is an important corollary to the skill of the invisible hand in guiding economic activity: When the government prevents prices from adjusting naturally to supply and demand, it impedes the invisible hand's ability to coordinate the millions of households and firms that make up the economy. This corollary explains why taxes adversely affect the allocation of resources: Taxes distort prices and thus the decisions of households and firms. It also explains the even greater harm caused by policies that directly control prices, such as rent control. And it explains the failure of communism. In communist countries, prices were not determined in the marketplace but were dictated by central planners. These planners lacked the information that gets reflected in prices when prices are free to respond to market

forces. Central planners failed because they tried to run the economy with one hand tied behind their backs—the invisible hand of the marketplace.

PRINCIPLE #7: GOVERNMENTS CAN SOMETIMES IMPROVE MARKET OUTCOMES

Although markets are usually a good way to organize economic activity, this rule has some important exceptions. There are two broad reasons for a government to intervene in the economy: to promote efficiency and to promote equity. That is, most policies aim either to enlarge the economic pie or to change how the pie is divided.

The invisible hand usually leads markets to allocate resources efficiently. Nonetheless, for various reasons, the invisible hand sometimes does not work. Economists use the term **market failure** to refer to a situation in which the market on its own fails to allocate resources efficiently.

One possible cause of market failure is an externality. An **externality** is the impact of one person's actions on the well-being of a bystander. The classic example of an external cost is pollution. If a chemical factory does not bear the entire cost of the smoke it emits, it will likely emit too much. Here, the government can raise economic well-being through environmental regulation. The classic example of an external benefit is the creation of knowledge. When a scientist makes an important discovery, he produces a valuable resource that other people can use. In this case, the government can raise economic well-being by subsidizing basic research, as in fact it does.

Another possible cause of market failure is market power. **Market power** refers to the ability of a single person (or small group of people) to unduly influence market prices. For example, suppose that everyone in town needs water but there is only one well. The owner of the well has market power—in this case a *monopoly*—over the sale of water. The well owner is not subject to the rigorous competition with which the invisible hand normally keeps self-interest in check. You will learn that, in this case, regulating the price that the monopolist charges can potentially enhance economic efficiency.

The invisible hand is even less able to ensure that economic prosperity is distributed fairly. A market economy rewards people according to their ability to produce things that other people are willing to pay for. The world's best basketball player earns more than the world's best chess player simply because people are willing to pay more to watch basketball than chess. The invisible hand does not ensure that everyone has sufficient food, decent clothing, and adequate health care. A goal of many public policies, such as the income tax and the welfare system, is to achieve a more equitable distribution of economic well-being.

To say that the government *can* improve on markets outcomes at times does not mean that it always *will*. Public policy is made not by angels but by a political process that is far from perfect. Sometimes policies are designed simply to reward the politically powerful. Sometimes they are made by well-intentioned leaders who are not fully informed. One goal of the study of economics is to help you judge when a government policy is justifiable to promote efficiency or equity and when it is not.

market failure

a situation in which a market left on its own fails to allocate resources efficiently

externality

the impact of one person's actions on the well-being of a bystander

market power

the ability of a single economic actor (or small group of actors) to have a substantial influence on market prices

QUICK QUIZ: List and briefly explain the three principles concerning economic interactions.

HOW THE ECONOMY AS A WHOLE WORKS

We started by discussing how individuals make decisions and then looked at how people interact with one another. All these decisions and interactions together make up “the economy.” The last three principles concern the workings of the economy as a whole.

PRINCIPLE #8: A COUNTRY’S STANDARD OF LIVING DEPENDS ON ITS ABILITY TO PRODUCE GOODS AND SERVICES

The differences in living standards around the world are staggering. In 1997 the average American had an income of about \$29,000. In the same year, the average Mexican earned \$8,000, and the average Nigerian earned \$900. Not surprisingly, this large variation in average income is reflected in various measures of the quality of life. Citizens of high-income countries have more TV sets, more cars, better nutrition, better health care, and longer life expectancy than citizens of low-income countries.

Changes in living standards over time are also large. In the United States, incomes have historically grown about 2 percent per year (after adjusting for changes in the cost of living). At this rate, average income doubles every 35 years. Over the past century, average income has risen about eightfold.

What explains these large differences in living standards among countries and over time? The answer is surprisingly simple. Almost all variation in living standards is attributable to differences in countries’ **productivity**—that is, the amount of goods and services produced from each hour of a worker’s time. In nations where workers can produce a large quantity of goods and services per unit of time, most people enjoy a high standard of living; in nations where workers are less productive, most people must endure a more meager existence. Similarly, the growth rate of a nation’s productivity determines the growth rate of its average income.

The fundamental relationship between productivity and living standards is simple, but its implications are far-reaching. If productivity is the primary determinant of living standards, other explanations must be of secondary importance. For example, it might be tempting to credit labor unions or minimum-wage laws for the rise in living standards of American workers over the past century. Yet the real hero of American workers is their rising productivity. As another example, some commentators have claimed that increased competition from Japan and other countries explains the slow growth in U.S. incomes over the past 30 years. Yet the real villain is not competition from abroad but flagging productivity growth in the United States.

The relationship between productivity and living standards also has profound implications for public policy. When thinking about how any policy will affect living standards, the key question is how it will affect our ability to produce goods and services. To boost living standards, policymakers need to raise productivity by ensuring that workers are well educated, have the tools needed to produce goods and services, and have access to the best available technology.

productivity

the amount of goods and services produced from each hour of a worker’s time

In the 1980s and 1990s, for example, much debate in the United States centered on the government's budget deficit—the excess of government spending over government revenue. As we will see, concern over the budget deficit was based largely on its adverse impact on productivity. When the government needs to finance a budget deficit, it does so by borrowing in financial markets, much as a student might borrow to finance a college education or a firm might borrow to finance a new factory. As the government borrows to finance its deficit, therefore, it reduces the quantity of funds available for other borrowers. The budget deficit thereby reduces investment both in human capital (the student's education) and physical capital (the firm's factory). Because lower investment today means lower productivity in the future, government budget deficits are generally thought to depress growth in living standards.

PRINCIPLE #9: PRICES RISE WHEN THE GOVERNMENT PRINTS TOO MUCH MONEY

In Germany in January 1921, a daily newspaper cost 0.30 marks. Less than two years later, in November 1922, the same newspaper cost 70,000,000 marks. All other prices in the economy rose by similar amounts. This episode is one of history's most spectacular examples of **inflation**, an increase in the overall level of prices in the economy.

Although the United States has never experienced inflation even close to that in Germany in the 1920s, inflation has at times been an economic problem. During the 1970s, for instance, the overall level of prices more than doubled, and President Gerald Ford called inflation “public enemy number one.” By contrast, inflation in the 1990s was about 3 percent per year; at this rate it would take more than

inflation

an increase in the overall level of prices in the economy



“Well it may have been 68 cents when you got in line, but it’s 74 cents now!”

20 years for prices to double. Because high inflation imposes various costs on society, keeping inflation at a low level is a goal of economic policymakers around the world.

What causes inflation? In almost all cases of large or persistent inflation, the culprit turns out to be the same—growth in the quantity of money. When a government creates large quantities of the nation's money, the value of the money falls. In Germany in the early 1920s, when prices were on average tripling every month, the quantity of money was also tripling every month. Although less dramatic, the economic history of the United States points to a similar conclusion: The high inflation of the 1970s was associated with rapid growth in the quantity of money, and the low inflation of the 1990s was associated with slow growth in the quantity of money.

PRINCIPLE #10: SOCIETY FACES A SHORT-RUN TRADEOFF BETWEEN INFLATION AND UNEMPLOYMENT

If inflation is so easy to explain, why do policymakers sometimes have trouble ridding the economy of it? One reason is that reducing inflation is often thought to cause a temporary rise in unemployment. The curve that illustrates this tradeoff between inflation and unemployment is called the **Phillips curve**, after the economist who first examined this relationship.

The Phillips curve remains a controversial topic among economists, but most economists today accept the idea that there is a short-run tradeoff between inflation and unemployment. This simply means that, over a period of a year or two, many economic policies push inflation and unemployment in opposite directions. Policymakers face this tradeoff regardless of whether inflation and unemployment both start out at high levels (as they were in the early 1980s), at low levels (as they were in the late 1990s), or someplace in between.

Why do we face this short-run tradeoff? According to a common explanation, it arises because some prices are slow to adjust. Suppose, for example, that the government reduces the quantity of money in the economy. In the long run, the only result of this policy change will be a fall in the overall level of prices. Yet not all prices will adjust immediately. It may take several years before all firms issue new catalogs, all unions make wage concessions, and all restaurants print new menus. That is, prices are said to be *sticky* in the short run.

Because prices are sticky, various types of government policy have short-run effects that differ from their long-run effects. When the government reduces the quantity of money, for instance, it reduces the amount that people spend. Lower spending, together with prices that are stuck too high, reduces the quantity of goods and services that firms sell. Lower sales, in turn, cause firms to lay off workers. Thus, the reduction in the quantity of money raises unemployment temporarily until prices have fully adjusted to the change.

The tradeoff between inflation and unemployment is only temporary, but it can last for several years. The Phillips curve is, therefore, crucial for understanding many developments in the economy. In particular, policymakers can exploit this tradeoff using various policy instruments. By changing the amount that the government spends, the amount it taxes, and the amount of money it prints, policymakers can, in the short run, influence the combination of inflation and unemployment that the economy experiences. Because these instruments of

Phillips curve

a curve that shows the short-run tradeoff between inflation and unemployment

monetary and fiscal policy are potentially so powerful, how policymakers should use these instruments to control the economy, if at all, is a subject of continuing debate.

QUICK QUIZ: List and briefly explain the three principles that describe how the economy as a whole works.

CONCLUSION

You now have a taste of what economics is all about. In the coming chapters we will develop many specific insights about people, markets, and economies. Mastering these insights will take some effort, but it is not an overwhelming task. The field of economics is based on a few basic ideas that can be applied in many different situations.

Throughout this book we will refer back to the *Ten Principles of Economics* highlighted in this chapter and summarized in Table 1-1. Whenever we do so, a building-blocks icon will be displayed in the margin, as it is now. But even when that icon is absent, you should keep these building blocks in mind. Even the most sophisticated economic analysis is built using the ten principles introduced here.



Table 1-1

TEN PRINCIPLES OF ECONOMICS

HOW PEOPLE MAKE DECISIONS	#1: People Face Tradeoffs
	#2: The Cost of Something Is What You Give Up to Get It
	#3: Rational People Think at the Margin
	#4: People Respond to Incentives
HOW PEOPLE INTERACT	#5: Trade Can Make Everyone Better Off
	#6: Markets Are Usually a Good Way to Organize Economic Activity
	#7: Governments Can Sometimes Improve Market Outcomes
HOW THE ECONOMY AS A WHOLE WORKS	#8: A Country's Standard of Living Depends on Its Ability to Produce Goods and Services
	#9: Prices Rise When the Government Prints Too Much Money
	#10: Society Faces a Short-Run Tradeoff between Inflation and Unemployment

Summary

- ◆ The fundamental lessons about individual decisionmaking are that people face tradeoffs among alternative goals, that the cost of any action is measured in terms of forgone opportunities, that rational people make decisions by comparing marginal costs and marginal benefits, and that people change their behavior in response to the incentives they face.
- ◆ The fundamental lessons about interactions among people are that trade can be mutually beneficial, that markets are usually a good way of coordinating trade among people, and that the government can potentially improve market outcomes if there is some market failure or if the market outcome is inequitable.
- ◆ The fundamental lessons about the economy as a whole are that productivity is the ultimate source of living standards, that money growth is the ultimate source of inflation, and that society faces a short-run tradeoff between inflation and unemployment.

Key Concepts

scarcity, p. 4
 economics, p. 4
 efficiency, p. 5
 equity, p. 5
 opportunity cost, p. 6

marginal changes, p. 6
 market economy, p. 9
 market failure, p. 11
 externality, p. 11
 market power, p. 11

productivity, p. 12
 inflation, p. 13
 Phillips curve, p. 14

Questions for Review

1. Give three examples of important tradeoffs that you face in your life.
2. What is the opportunity cost of seeing a movie?
3. Water is necessary for life. Is the marginal benefit of a glass of water large or small?
4. Why should policymakers think about incentives?
5. Why isn't trade among countries like a game with some winners and some losers?
6. What does the "invisible hand" of the marketplace do?
7. Explain the two main causes of market failure and give an example of each.
8. Why is productivity important?
9. What is inflation, and what causes it?
10. How are inflation and unemployment related in the short run?

Problems and Applications

1. Describe some of the tradeoffs faced by the following:
 - a. a family deciding whether to buy a new car
 - b. a member of Congress deciding how much to spend on national parks
 - c. a company president deciding whether to open a new factory
 - d. a professor deciding how much to prepare for class
2. You are trying to decide whether to take a vacation. Most of the costs of the vacation (airfare, hotel, forgone wages) are measured in dollars, but the benefits of the vacation are psychological. How can you compare the benefits to the costs?
3. You were planning to spend Saturday working at your part-time job, but a friend asks you to go skiing. What is the true cost of going skiing? Now suppose that you had been planning to spend the day studying at the library. What is the cost of going skiing in this case? Explain.
4. You win \$100 in a basketball pool. You have a choice between spending the money now or putting it away for a year in a bank account that pays 5 percent interest. What is the opportunity cost of spending the \$100 now?
5. The company that you manage has invested \$5 million in developing a new product, but the development is not quite finished. At a recent meeting, your salespeople report that the introduction of competing products has reduced the expected sales of your new product to \$3 million. If it would cost \$1 million to finish

development and make the product, should you go ahead and do so? What is the most that you should pay to complete development?

6. Three managers of the Magic Potion Company are discussing a possible increase in production. Each suggests a way to make this decision.

HARRY: We should examine whether our company's productivity—gallons of potion per worker—would rise or fall.

RON: We should examine whether our average cost—cost per worker—would rise or fall.

HERMIONE: We should examine whether the extra revenue from selling the additional potion would be greater or smaller than the extra costs.

Who do you think is right? Why?

7. The Social Security system provides income for people over age 65. If a recipient of Social Security decides to work and earn some income, the amount he or she receives in Social Security benefits is typically reduced.
 - a. How does the provision of Social Security affect people's incentive to save while working?
 - b. How does the reduction in benefits associated with higher earnings affect people's incentive to work past age 65?
8. A recent bill reforming the government's antipoverty programs limited many welfare recipients to only two years of benefits.
 - a. How does this change affect the incentives for working?
 - b. How might this change represent a tradeoff between equity and efficiency?
9. Your roommate is a better cook than you are, but you can clean more quickly than your roommate can. If your roommate did all of the cooking and you did all of the cleaning, would your chores take you more or less time than if you divided each task evenly? Give a similar example of how specialization and trade can make two countries both better off.
10. Suppose the United States adopted central planning for its economy, and you became the chief planner. Among the millions of decisions that you need to make for next year are how many compact discs to produce, what artists to record, and who should receive the discs.
 - a. To make these decisions intelligently, what information would you need about the compact disc industry? What information would you need about each of the people in the United States?
 - b. How would your decisions about CDs affect some of your other decisions, such as how many CD players to make or cassette tapes to produce? How might some of your other decisions about the economy change your views about CDs?
11. Explain whether each of the following government activities is motivated by a concern about equity or a concern about efficiency. In the case of efficiency, discuss the type of market failure involved.
 - a. regulating cable-TV prices
 - b. providing some poor people with vouchers that can be used to buy food
 - c. prohibiting smoking in public places
 - d. breaking up Standard Oil (which once owned 90 percent of all oil refineries) into several smaller companies
 - e. imposing higher personal income tax rates on people with higher incomes
 - f. instituting laws against driving while intoxicated
12. Discuss each of the following statements from the standpoints of equity and efficiency.
 - a. "Everyone in society should be guaranteed the best health care possible."
 - b. "When workers are laid off, they should be able to collect unemployment benefits until they find a new job."
13. In what ways is your standard of living different from that of your parents or grandparents when they were your age? Why have these changes occurred?
14. Suppose Americans decide to save more of their incomes. If banks lend this extra saving to businesses, which use the funds to build new factories, how might this lead to faster growth in productivity? Who do you suppose benefits from the higher productivity? Is society getting a free lunch?
15. Suppose that when everyone wakes up tomorrow, they discover that the government has given them an additional amount of money equal to the amount they already had. Explain what effect this doubling of the money supply will likely have on the following:
 - a. the total amount spent on goods and services
 - b. the quantity of goods and services purchased if prices are sticky
 - c. the prices of goods and services if prices can adjust
16. Imagine that you are a policymaker trying to decide whether to reduce the rate of inflation. To make an intelligent decision, what would you need to know about inflation, unemployment, and the tradeoff between them?

2



THINKING LIKE AN ECONOMIST

Every field of study has its own language and its own way of thinking. Mathematicians talk about axioms, integrals, and vector spaces. Psychologists talk about ego, id, and cognitive dissonance. Lawyers talk about venue, torts, and promissory estoppel.

Economics is no different. Supply, demand, elasticity, comparative advantage, consumer surplus, deadweight loss—these terms are part of the economist's language. In the coming chapters, you will encounter many new terms and some familiar words that economists use in specialized ways. At first, this new language may seem needlessly arcane. But, as you will see, its value lies in its ability to provide you a new and useful way of thinking about the world in which you live.

The single most important purpose of this book is to help you learn the economist's way of thinking. Of course, just as you cannot become a mathematician, psychologist, or lawyer overnight, learning to think like an economist will take

IN THIS CHAPTER YOU WILL . . .

*See how economists
apply the methods
of science*

*Consider how
assumptions and
models can shed
light on the world*

*Learn two simple
models—the circular
flow and the
production
possibilities frontier*

*Distinguish between
microeconomics and
macroeconomics*

*Learn the difference
between positive and
normative statements*

*Examine the role of
economists in
making policy*

*Consider why
economists
sometimes disagree
with one another*

some time. Yet with a combination of theory, case studies, and examples of economics in the news, this book will give you ample opportunity to develop and practice this skill.

Before delving into the substance and details of economics, it is helpful to have an overview of how economists approach the world. This chapter, therefore, discusses the field's methodology. What is distinctive about how economists confront a question? What does it mean to think like an economist?

THE ECONOMIST AS SCIENTIST

Economists try to address their subject with a scientist's objectivity. They approach the study of the economy in much the same way as a physicist approaches the study of matter and a biologist approaches the study of life: They devise theories, collect data, and then analyze these data in an attempt to verify or refute their theories.

To beginners, it can seem odd to claim that economics is a science. After all, economists do not work with test tubes or telescopes. The essence of science,



"I'm a social scientist, Michael. That means I can't explain electricity or anything like that, but if you ever want to know about people I'm your man."

however, is the *scientific method*—the dispassionate development and testing of theories about how the world works. This method of inquiry is as applicable to studying a nation's economy as it is to studying the earth's gravity or a species' evolution. As Albert Einstein once put it, "The whole of science is nothing more than the refinement of everyday thinking."

Although Einstein's comment is as true for social sciences such as economics as it is for natural sciences such as physics, most people are not accustomed to looking at society through the eyes of a scientist. Let's therefore discuss some of the ways in which economists apply the logic of science to examine how an economy works.

THE SCIENTIFIC METHOD: OBSERVATION, THEORY, AND MORE OBSERVATION

Isaac Newton, the famous seventeenth-century scientist and mathematician, allegedly became intrigued one day when he saw an apple fall from an apple tree. This observation motivated Newton to develop a theory of gravity that applies not only to an apple falling to the earth but to any two objects in the universe. Subsequent testing of Newton's theory has shown that it works well in many circumstances (although, as Einstein would later emphasize, not in all circumstances). Because Newton's theory has been so successful at explaining observation, it is still taught today in undergraduate physics courses around the world.

This interplay between theory and observation also occurs in the field of economics. An economist might live in a country experiencing rapid increases in prices and be moved by this observation to develop a theory of inflation. The theory might assert that high inflation arises when the government prints too much money. (As you may recall, this was one of the *Ten Principles of Economics* in Chapter 1.) To test this theory, the economist could collect and analyze data on prices and money from many different countries. If growth in the quantity of money were not at all related to the rate at which prices are rising, the economist would start to doubt the validity of his theory of inflation. If money growth and inflation were strongly correlated in international data, as in fact they are, the economist would become more confident in his theory.

Although economists use theory and observation like other scientists, they do face an obstacle that makes their task especially challenging: Experiments are often difficult in economics. Physicists studying gravity can drop many objects in their laboratories to generate data to test their theories. By contrast, economists studying inflation are not allowed to manipulate a nation's monetary policy simply to generate useful data. Economists, like astronomers and evolutionary biologists, usually have to make do with whatever data the world happens to give them.

To find a substitute for laboratory experiments, economists pay close attention to the natural experiments offered by history. When a war in the Middle East interrupts the flow of crude oil, for instance, oil prices skyrocket around the world. For consumers of oil and oil products, such an event depresses living standards. For economic policymakers, it poses a difficult choice about how best to respond. But for economic scientists, it provides an opportunity to study the effects of a key natural resource on the world's economies, and this opportunity persists long after the wartime increase in oil prices is over. Throughout this book, therefore, we consider many historical episodes. These episodes are valuable to study because they



give us insight into the economy of the past and, more important, because they allow us to illustrate and evaluate economic theories of the present.

THE ROLE OF ASSUMPTIONS

If you ask a physicist how long it would take for a marble to fall from the top of a ten-story building, she will answer the question by assuming that the marble falls in a vacuum. Of course, this assumption is false. In fact, the building is surrounded by air, which exerts friction on the falling marble and slows it down. Yet the physicist will correctly point out that friction on the marble is so small that its effect is negligible. Assuming the marble falls in a vacuum greatly simplifies the problem without substantially affecting the answer.

Economists make assumptions for the same reason: Assumptions can make the world easier to understand. To study the effects of international trade, for example, we may assume that the world consists of only two countries and that each country produces only two goods. Of course, the real world consists of dozens of countries, each of which produces thousands of different types of goods. But by assuming two countries and two goods, we can focus our thinking. Once we understand international trade in an imaginary world with two countries and two goods, we are in a better position to understand international trade in the more complex world in which we live.

The art in scientific thinking—whether in physics, biology, or economics—is deciding which assumptions to make. Suppose, for instance, that we were dropping a beach ball rather than a marble from the top of the building. Our physicist would realize that the assumption of no friction is far less accurate in this case: Friction exerts a greater force on a beach ball than on a marble. The assumption that gravity works in a vacuum is reasonable for studying a falling marble but not for studying a falling beach ball.

Similarly, economists use different assumptions to answer different questions. Suppose that we want to study what happens to the economy when the government changes the number of dollars in circulation. An important piece of this analysis, it turns out, is how prices respond. Many prices in the economy change infrequently; the newsstand prices of magazines, for instance, are changed only every few years. Knowing this fact may lead us to make different assumptions when studying the effects of the policy change over different time horizons. For studying the short-run effects of the policy, we may assume that prices do not change much. We may even make the extreme and artificial assumption that all prices are completely fixed. For studying the long-run effects of the policy, however, we may assume that all prices are completely flexible. Just as a physicist uses different assumptions when studying falling marbles and falling beach balls, economists use different assumptions when studying the short-run and long-run effects of a change in the quantity of money.

ECONOMIC MODELS

High school biology teachers teach basic anatomy with plastic replicas of the human body. These models have all the major organs—the heart, the liver, the kidneys, and so on. The models allow teachers to show their students in a simple way how the important parts of the body fit together. Of course, these plastic models

are not actual human bodies, and no one would mistake the model for a real person. These models are stylized, and they omit many details. Yet despite this lack of realism—indeed, because of this lack of realism—studying these models is useful for learning how the human body works.

Economists also use models to learn about the world, but instead of being made of plastic, they are most often composed of diagrams and equations. Like a biology teacher's plastic model, economic models omit many details to allow us to see what is truly important. Just as the biology teacher's model does not include all of the body's muscles and capillaries, an economist's model does not include every feature of the economy.

As we use models to examine various economic issues throughout this book, you will see that all the models are built with assumptions. Just as a physicist begins the analysis of a falling marble by assuming away the existence of friction, economists assume away many of the details of the economy that are irrelevant for studying the question at hand. All models—in physics, biology, or economics—simplify reality in order to improve our understanding of it.

OUR FIRST MODEL: THE CIRCULAR-FLOW DIAGRAM

The economy consists of millions of people engaged in many activities—buying, selling, working, hiring, manufacturing, and so on. To understand how the economy works, we must find some way to simplify our thinking about all these activities. In other words, we need a model that explains, in general terms, how the economy is organized and how participants in the economy interact with one another.

Figure 2-1 presents a visual model of the economy, called a **circular-flow diagram**. In this model, the economy has two types of decisionmakers—households and firms. Firms produce goods and services using inputs, such as labor, land, and capital (buildings and machines). These inputs are called the *factors of production*. Households own the factors of production and consume all the goods and services that the firms produce.

Households and firms interact in two types of markets. In the *markets for goods and services*, households are buyers and firms are sellers. In particular, households buy the output of goods and services that firms produce. In the *markets for the factors of production*, households are sellers and firms are buyers. In these markets, households provide firms the inputs that the firms use to produce goods and services. The circular-flow diagram offers a simple way of organizing all the economic transactions that occur between households and firms in the economy.

The inner loop of the circular-flow diagram represents the flows of goods and services between households and firms. The households sell the use of their labor, land, and capital to the firms in the markets for the factors of production. The firms then use these factors to produce goods and services, which in turn are sold to households in the markets for goods and services. Hence, the factors of production flow from households to firms, and goods and services flow from firms to households.

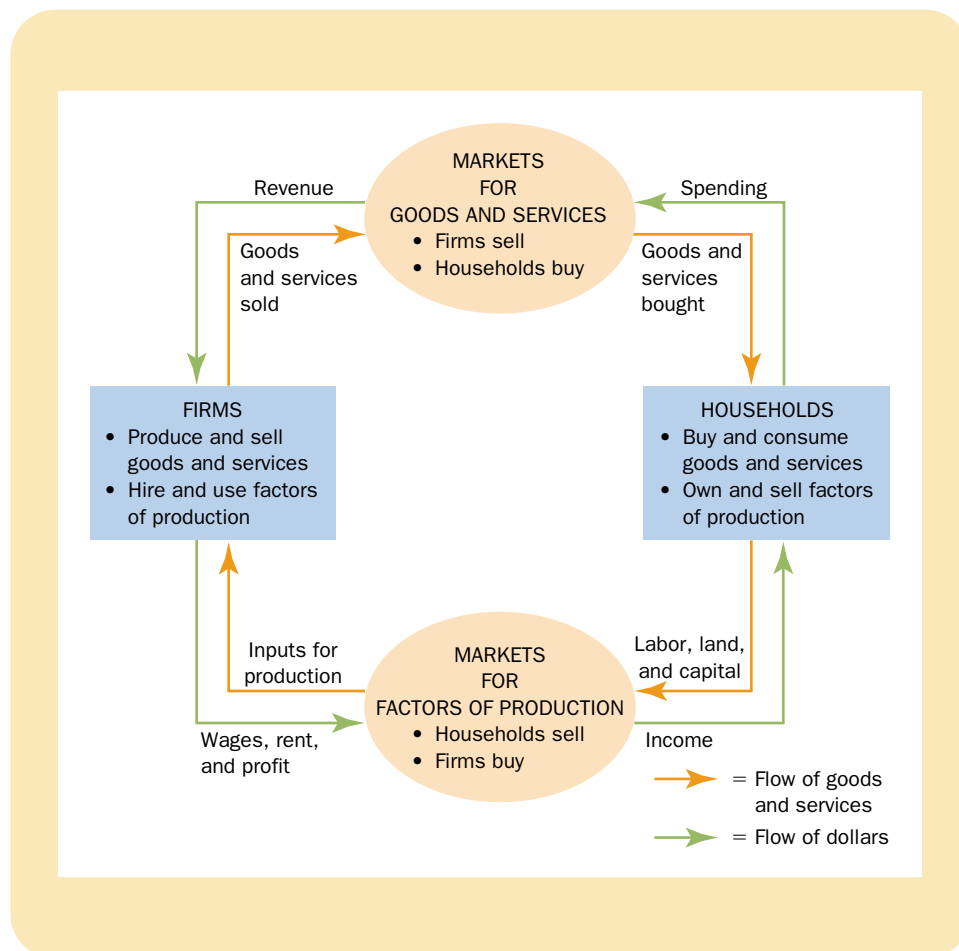
The outer loop of the circular-flow diagram represents the corresponding flow of dollars. The households spend money to buy goods and services from the firms. The firms use some of the revenue from these sales to pay for the factors of

circular-flow diagram

a visual model of the economy that shows how dollars flow through markets among households and firms

Figure 2-1

THE CIRCULAR FLOW. This diagram is a schematic representation of the organization of the economy. Decisions are made by households and firms. Households and firms interact in the markets for goods and services (where households are buyers and firms are sellers) and in the markets for the factors of production (where firms are buyers and households are sellers). The outer set of arrows shows the flow of dollars, and the inner set of arrows shows the corresponding flow of goods and services.



production, such as the wages of their workers. What's left is the profit of the firm owners, who themselves are members of households. Hence, spending on goods and services flows from households to firms, and income in the form of wages, rent, and profit flows from firms to households.

Let's take a tour of the circular flow by following a dollar bill as it makes its way from person to person through the economy. Imagine that the dollar begins at a household, sitting in, say, your wallet. If you want to buy a cup of coffee, you take the dollar to one of the economy's markets for goods and services, such as your local Starbucks coffee shop. There you spend it on your favorite drink. When the dollar moves into the Starbucks cash register, it becomes revenue for the firm. The dollar doesn't stay at Starbucks for long, however, because the firm uses it to buy inputs in the markets for the factors of production. For instance, Starbucks might use the dollar to pay rent to its landlord for the space it occupies or to pay the wages of its workers. In either case, the dollar enters the income of some household and, once again, is back in someone's wallet. At that point, the story of the economy's circular flow starts once again.

The circular-flow diagram in Figure 2-1 is one simple model of the economy. It dispenses with details that, for some purposes, are significant. A more complex

and realistic circular-flow model would include, for instance, the roles of government and international trade. Yet these details are not crucial for a basic understanding of how the economy is organized. Because of its simplicity, this circular-flow diagram is useful to keep in mind when thinking about how the pieces of the economy fit together.

OUR SECOND MODEL: THE PRODUCTION POSSIBILITIES FRONTIER

Most economic models, unlike the circular-flow diagram, are built using the tools of mathematics. Here we consider one of the simplest such models, called the production possibilities frontier, and see how this model illustrates some basic economic ideas.

Although real economies produce thousands of goods and services, let's imagine an economy that produces only two goods—cars and computers. Together the car industry and the computer industry use all of the economy's factors of production. The **production possibilities frontier** is a graph that shows the various combinations of output—in this case, cars and computers—that the economy can possibly produce given the available factors of production and the available production technology that firms can use to turn these factors into output.

Figure 2-2 is an example of a production possibilities frontier. In this economy, if all resources were used in the car industry, the economy would produce 1,000 cars and no computers. If all resources were used in the computer industry, the economy would produce 3,000 computers and no cars. The two end points of the production possibilities frontier represent these extreme possibilities. If the

production possibilities frontier

a graph that shows the combinations of output that the economy can possibly produce given the available factors of production and the available production technology

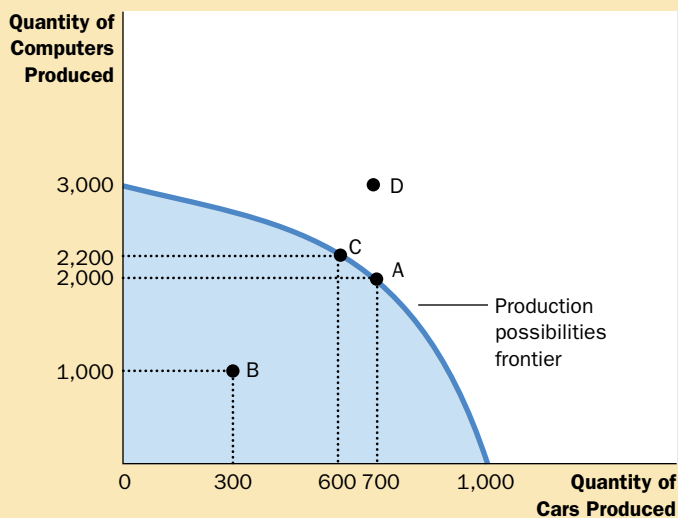


Figure 2-2

THE PRODUCTION POSSIBILITIES FRONTIER. The production possibilities frontier shows the combinations of output—in this case, cars and computers—that the economy can possibly produce. The economy can produce any combination on or inside the frontier. Points outside the frontier are not feasible given the economy's resources.

economy were to divide its resources between the two industries, it could produce 700 cars and 2,000 computers, shown in the figure by point A. By contrast, the outcome at point D is not possible because resources are scarce: The economy does not have enough of the factors of production to support that level of output. In other words, the economy can produce at any point on or inside the production possibilities frontier, but it cannot produce at points outside the frontier.

An outcome is said to be *efficient* if the economy is getting all it can from the scarce resources it has available. Points on (rather than inside) the production possibilities frontier represent efficient levels of production. When the economy is producing at such a point, say point A, there is no way to produce more of one good without producing less of the other. Point B represents an *inefficient* outcome. For some reason, perhaps widespread unemployment, the economy is producing less than it could from the resources it has available: It is producing only 300 cars and 1,000 computers. If the source of the inefficiency were eliminated, the economy could move from point B to point A, increasing production of both cars (to 700) and computers (to 2,000).



One of the *Ten Principles of Economics* discussed in Chapter 1 is that people face tradeoffs. The production possibilities frontier shows one tradeoff that society faces. Once we have reached the efficient points on the frontier, the only way of getting more of one good is to get less of the other. When the economy moves from point A to point C, for instance, society produces more computers but at the expense of producing fewer cars.

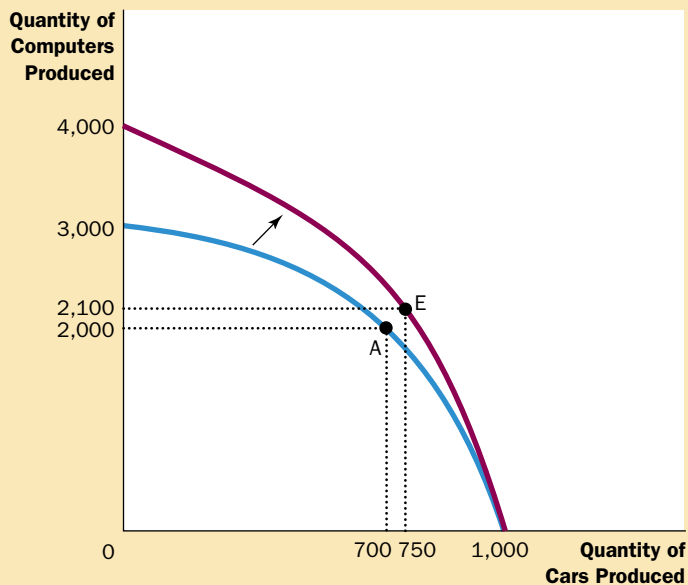


Another of the *Ten Principles of Economics* is that the cost of something is what you give up to get it. This is called the *opportunity cost*. The production possibilities frontier shows the opportunity cost of one good as measured in terms of the other good. When society reallocates some of the factors of production from the car industry to the computer industry, moving the economy from point A to point C, it gives up 100 cars to get 200 additional computers. In other words, when the economy is at point A, the opportunity cost of 200 computers is 100 cars.

Notice that the production possibilities frontier in Figure 2-2 is bowed outward. This means that the opportunity cost of cars in terms of computers depends on how much of each good the economy is producing. When the economy is using most of its resources to make cars, the production possibilities frontier is quite steep. Because even workers and machines best suited to making computers are being used to make cars, the economy gets a substantial increase in the number of computers for each car it gives up. By contrast, when the economy is using most of its resources to make computers, the production possibilities frontier is quite flat. In this case, the resources best suited to making computers are already in the computer industry, and each car the economy gives up yields only a small increase in the number of computers.

The production possibilities frontier shows the tradeoff between the production of different goods at a given time, but the tradeoff can change over time. For example, if a technological advance in the computer industry raises the number of computers that a worker can produce per week, the economy can make more computers for any given number of cars. As a result, the production possibilities frontier shifts outward, as in Figure 2-3. Because of this economic growth, society might move production from point A to point E, enjoying more computers and more cars.

The production possibilities frontier simplifies a complex economy to highlight and clarify some basic ideas. We have used it to illustrate some of the

**Figure 2-3**

A SHIFT IN THE PRODUCTION POSSIBILITIES FRONTIER. An economic advance in the computer industry shifts the production possibilities frontier outward, increasing the number of cars and computers the economy can produce.

concepts mentioned briefly in Chapter 1: scarcity, efficiency, tradeoffs, opportunity cost, and economic growth. As you study economics, these ideas will recur in various forms. The production possibilities frontier offers one simple way of thinking about them.

MICROECONOMICS AND MACROECONOMICS

Many subjects are studied on various levels. Consider biology, for example. Molecular biologists study the chemical compounds that make up living things. Cellular biologists study cells, which are made up of many chemical compounds and, at the same time, are themselves the building blocks of living organisms. Evolutionary biologists study the many varieties of animals and plants and how species change gradually over the centuries.

Economics is also studied on various levels. We can study the decisions of individual households and firms. Or we can study the interaction of households and firms in markets for specific goods and services. Or we can study the operation of the economy as a whole, which is just the sum of the activities of all these decision-makers in all these markets.

The field of economics is traditionally divided into two broad subfields. **Microeconomics** is the study of how households and firms make decisions and how they interact in specific markets. **Macroeconomics** is the study of economy-wide phenomena. A microeconomist might study the effects of rent control on housing in New York City, the impact of foreign competition on the U.S. auto industry, or the effects of compulsory school attendance on workers' earnings. A

microeconomics

the study of how households and firms make decisions and how they interact in markets

macroeconomics

the study of economy-wide phenomena, including inflation, unemployment, and economic growth

macroeconomist might study the effects of borrowing by the federal government, the changes over time in the economy's rate of unemployment, or alternative policies to raise growth in national living standards.

Microeconomics and macroeconomics are closely intertwined. Because changes in the overall economy arise from the decisions of millions of individuals, it is impossible to understand macroeconomic developments without considering the associated microeconomic decisions. For example, a macroeconomist might study the effect of a cut in the federal income tax on the overall production of goods and services. To analyze this issue, he or she must consider how the tax cut affects the decisions of households about how much to spend on goods and services.

Despite the inherent link between microeconomics and macroeconomics, the two fields are distinct. In economics, as in biology, it may seem natural to begin with the smallest unit and build up. Yet doing so is neither necessary nor always the best way to proceed. Evolutionary biology is, in a sense, built upon molecular biology, since species are made up of molecules. Yet molecular biology and evolutionary biology are separate fields, each with its own questions and its own methods. Similarly, because microeconomics and macroeconomics address different questions, they sometimes take quite different approaches and are often taught in separate courses.

QUICK QUIZ: In what sense is economics like a science? ♦ Draw a production possibilities frontier for a society that produces food and clothing. Show an efficient point, an inefficient point, and an infeasible point. Show the effects of a drought. ♦ Define *microeconomics* and *macroeconomics*.

THE ECONOMIST AS POLICY ADVISER

Often economists are asked to explain the causes of economic events. Why, for example, is unemployment higher for teenagers than for older workers? Sometimes economists are asked to recommend policies to improve economic outcomes. What, for instance, should the government do to improve the economic well-being of teenagers? When economists are trying to explain the world, they are scientists. When they are trying to help improve it, they are policy advisers.

POSITIVE VERSUS NORMATIVE ANALYSIS

To help clarify the two roles that economists play, we begin by examining the use of language. Because scientists and policy advisers have different goals, they use language in different ways.

For example, suppose that two people are discussing minimum-wage laws. Here are two statements you might hear:

POLLY: Minimum-wage laws cause unemployment.

NORMA: The government should raise the minimum wage.

Ignoring for now whether you agree with these statements, notice that Polly and Norma differ in what they are trying to do. Polly is speaking like a scientist: She is making a claim about how the world works. Norma is speaking like a policy adviser: She is making a claim about how she would like to change the world.

In general, statements about the world are of two types. One type, such as Polly's, is positive. **Positive statements** are descriptive. They make a claim about how the world *is*. A second type of statement, such as Norma's, is normative. **Normative statements** are prescriptive. They make a claim about how the world *ought to be*.

A key difference between positive and normative statements is how we judge their validity. We can, in principle, confirm or refute positive statements by examining evidence. An economist might evaluate Polly's statement by analyzing data on changes in minimum wages and changes in unemployment over time. By contrast, evaluating normative statements involves values as well as facts. Norma's statement cannot be judged using data alone. Deciding what is good or bad policy is not merely a matter of science. It also involves our views on ethics, religion, and political philosophy.

Of course, positive and normative statements may be related. Our positive views about how the world works affect our normative views about what policies are desirable. Polly's claim that the minimum wage causes unemployment, if true, might lead us to reject Norma's conclusion that the government should raise the minimum wage. Yet our normative conclusions cannot come from positive analysis alone. Instead, they require both positive analysis and value judgments.

As you study economics, keep in mind the distinction between positive and normative statements. Much of economics just tries to explain how the economy works. Yet often the goal of economics is to improve how the economy works. When you hear economists making normative statements, you know they have crossed the line from scientist to policy adviser.

positive statements

claims that attempt to describe the world as it is

normative statements

claims that attempt to prescribe how the world should be

ECONOMISTS IN WASHINGTON

President Harry Truman once said that he wanted to find a one-armed economist. When he asked his economists for advice, they always answered, "On the one hand, . . . On the other hand, . . ."

Truman was right in realizing that economists' advice is not always straightforward. This tendency is rooted in one of the *Ten Principles of Economics* in Chapter 1: People face tradeoffs. Economists are aware that tradeoffs are involved in most policy decisions. A policy might increase efficiency at the cost of equity. It might help future generations but hurt current generations. An economist who says that all policy decisions are easy is an economist not to be trusted.

Truman was also not alone among presidents in relying on the advice of economists. Since 1946, the president of the United States has received guidance from the Council of Economic Advisers, which consists of three members and a staff of several dozen economists. The council, whose offices are just a few steps from the White House, has no duty other than to advise the president and to write the annual *Economic Report of the President*.

The president also receives input from economists in many administrative departments. Economists at the Department of Treasury help design tax policy. Economists at the Department of Labor analyze data on workers and those looking for





“Let’s switch. I’ll make the policy, you implement it, and he’ll explain it.”

work in order to help formulate labor-market policies. Economists at the Department of Justice help enforce the nation’s antitrust laws.

Economists are also found outside the administrative branch of government. To obtain independent evaluations of policy proposals, Congress relies on the advice of the Congressional Budget Office, which is staffed by economists. The Federal Reserve, the quasi-governmental institution that sets the nation’s monetary policy, employs hundreds of economists to analyze economic developments in the United States and throughout the world. Table 2-1 lists the Web sites of some of these agencies.

The influence of economists on policy goes beyond their role as advisers: Their research and writings often affect policy indirectly. Economist John Maynard Keynes offered this observation:

The ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed, the world is ruled by little else. Practical men, who believe themselves to be quite exempt from intellectual influences, are usually the slaves of some defunct economist. Madmen in authority, who hear voices in the air, are distilling their frenzy from some academic scribbler of a few years back.

Table 2-1

WEB SITES. Here are the Web sites for a few of the government agencies that are responsible for collecting economic data and making economic policy.

Department of Commerce
Bureau of Labor Statistics
Congressional Budget Office
Federal Reserve Board

www.doc.gov
www.bls.gov
www.cbo.gov
www.federalreserve.gov

Although these words were written in 1935, they remain true today. Indeed, the “academic scribbler” now influencing public policy is often Keynes himself.

QUICK QUIZ: Give an example of a positive statement and an example of a normative statement. ♦ Name three parts of government that regularly rely on advice from economists.

WHY ECONOMISTS DISAGREE

“If all economists were laid end to end, they would not reach a conclusion.” This quip from George Bernard Shaw is revealing. Economists as a group are often criticized for giving conflicting advice to policymakers. President Ronald Reagan once joked that if the game Trivial Pursuit were designed for economists, it would have 100 questions and 3,000 answers.

Why do economists so often appear to give conflicting advice to policymakers? There are two basic reasons:

- ♦ Economists may disagree about the validity of alternative positive theories about how the world works.
- ♦ Economists may have different values and, therefore, different normative views about what policy should try to accomplish.

Let’s discuss each of these reasons.

DIFFERENCES IN SCIENTIFIC JUDGMENTS

Several centuries ago, astronomers debated whether the earth or the sun was at the center of the solar system. More recently, meteorologists have debated whether the earth is experiencing “global warming” and, if so, why. Science is a search for understanding about the world around us. It is not surprising that as the search continues, scientists can disagree about the direction in which truth lies.

Economists often disagree for the same reason. Economics is a young science, and there is still much to be learned. Economists sometimes disagree because they have different hunches about the validity of alternative theories or about the size of important parameters.

For example, economists disagree about whether the government should levy taxes based on a household’s income or its consumption (spending). Advocates of a switch from the current income tax to a consumption tax believe that the change would encourage households to save more, because income that is saved would not be taxed. Higher saving, in turn, would lead to more rapid growth in productivity and living standards. Advocates of the current income tax believe that household saving would not respond much to a change in the tax laws. These two groups of economists hold different normative views about the tax system because they have different positive views about the responsiveness of saving to tax incentives.

DIFFERENCES IN VALUES

Suppose that Peter and Paul both take the same amount of water from the town well. To pay for maintaining the well, the town taxes its residents. Peter has income of \$50,000 and is taxed \$5,000, or 10 percent of his income. Paul has income of \$10,000 and is taxed \$2,000, or 20 percent of his income.

Is this policy fair? If not, who pays too much and who pays too little? Does it matter whether Paul's low income is due to a medical disability or to his decision to pursue a career in acting? Does it matter whether Peter's high income is due to a large inheritance or to his willingness to work long hours at a dreary job?

These are difficult questions on which people are likely to disagree. If the town hired two experts to study how the town should tax its residents to pay for the well, we would not be surprised if they offered conflicting advice.

This simple example shows why economists sometimes disagree about public policy. As we learned earlier in our discussion of normative and positive analysis, policies cannot be judged on scientific grounds alone. Economists give conflicting advice sometimes because they have different values. Perfecting the science of economics will not tell us whether it is Peter or Paul who pays too much.

PERCEPTION VERSUS REALITY

Because of differences in scientific judgments and differences in values, some disagreement among economists is inevitable. Yet one should not overstate the amount of disagreement. In many cases, economists do offer a united view.

Table 2-2 contains ten propositions about economic policy. In a survey of economists in business, government, and academia, these propositions were endorsed by an overwhelming majority of respondents. Most of these propositions would fail to command a similar consensus among the general public.

The first proposition in the table is about rent control. For reasons we will discuss in Chapter 6, almost all economists believe that rent control adversely affects the availability and quality of housing and is a very costly way of helping the most needy members of society. Nonetheless, many city governments choose to ignore the advice of economists and place ceilings on the rents that landlords may charge their tenants.

The second proposition in the table concerns tariffs and import quotas. For reasons we will discuss in Chapter 3 and more fully in Chapter 9, almost all economists oppose such barriers to free trade. Nonetheless, over the years, the president and Congress have chosen to restrict the import of certain goods. In 1993 the North American Free Trade Agreement (NAFTA), which reduced barriers to trade among the United States, Canada, and Mexico, passed Congress, but only by a narrow margin, despite overwhelming support from economists. In this case, economists did offer united advice, but many members of Congress chose to ignore it.

Why do policies such as rent control and import quotas persist if the experts are united in their opposition? The reason may be that economists have not yet convinced the general public that these policies are undesirable. One purpose of this book is to make you understand the economist's view of these and other subjects and, perhaps, to persuade you that it is the right one.

Table 2-2

TEN PROPOSITIONS ABOUT
WHICH MOST ECONOMISTS
AGREE

PROPOSITION (AND PERCENTAGE OF ECONOMISTS WHO AGREE)

1. A ceiling on rents reduces the quantity and quality of housing available. (93%)
2. Tariffs and import quotas usually reduce general economic welfare. (93%)
3. Flexible and floating exchange rates offer an effective international monetary arrangement. (90%)
4. Fiscal policy (e.g., tax cut and/or government expenditure increase) has a significant stimulative impact on a less than fully employed economy. (90%)
5. If the federal budget is to be balanced, it should be done over the business cycle rather than yearly. (85%)
6. Cash payments increase the welfare of recipients to a greater degree than do transfers-in-kind of equal cash value. (84%)
7. A large federal budget deficit has an adverse effect on the economy. (83%)
8. A minimum wage increases unemployment among young and unskilled workers. (79%)
9. The government should restructure the welfare system along the lines of a "negative income tax." (79%)
10. Effluent taxes and marketable pollution permits represent a better approach to pollution control than imposition of pollution ceilings. (78%)

SOURCE: Richard M. Alston, J. R. Kearl, and Michael B. Vaughn, "Is There Consensus among Economists in the 1990s?" *American Economic Review* (May 1992): 203–209.

QUICK QUIZ: Why might economic advisers to the president disagree about a question of policy?

LET'S GET GOING

The first two chapters of this book have introduced you to the ideas and methods of economics. We are now ready to get to work. In the next chapter we start learning in more detail the principles of economic behavior and economic policy.

As you proceed through this book, you will be asked to draw on many of your intellectual skills. You might find it helpful to keep in mind some advice from the great economist John Maynard Keynes:

The study of economics does not seem to require any specialized gifts of an unusually high order. Is it not . . . a very easy subject compared with the higher branches of philosophy or pure science? An easy subject, at which very few excel! The paradox finds its explanation, perhaps, in that the master-economist must possess a rare *combination* of gifts. He must be mathematician, historian, statesman, philosopher—in some degree. He must understand symbols and speak in words. He must contemplate the particular in terms of the general, and touch abstract and concrete in the same flight of thought. He must study the

present in the light of the past for the purposes of the future. No part of man's nature or his institutions must lie entirely outside his regard. He must be purposeful and disinterested in a simultaneous mood; as aloof and incorruptible as an artist, yet sometimes as near the earth as a politician.

It is a tall order. But with practice, you will become more and more accustomed to thinking like an economist.

Summary

- ◆ Economists try to address their subject with a scientist's objectivity. Like all scientists, they make appropriate assumptions and build simplified models in order to understand the world around them. Two simple economic models are the circular-flow diagram and the production possibilities frontier.
- ◆ The field of economics is divided into two subfields: microeconomics and macroeconomics. Microeconomists study decisionmaking by households and firms and the interaction among households and firms in the marketplace. Macroeconomists study the forces and trends that affect the economy as a whole.
- ◆ A positive statement is an assertion about how the world *is*. A normative statement is an assertion about how the world *ought to be*. When economists make normative statements, they are acting more as policy advisers than scientists.
- ◆ Economists who advise policymakers offer conflicting advice either because of differences in scientific judgments or because of differences in values. At other times, economists are united in the advice they offer, but policymakers may choose to ignore it.

Key Concepts

circular-flow diagram, p. 23

production possibilities frontier, p. 25

microeconomics, p. 27

macroeconomics, p. 27

positive statements, p. 29

normative statements, p. 29

Questions for Review

1. How is economics like a science?
2. Why do economists make assumptions?
3. Should an economic model describe reality exactly?
4. Draw and explain a production possibilities frontier for an economy that produces milk and cookies. What happens to this frontier if disease kills half of the economy's cow population?
5. Use a production possibilities frontier to describe the idea of "efficiency."
6. What are the two subfields into which economics is divided? Explain what each subfield studies.
7. What is the difference between a positive and a normative statement? Give an example of each.
8. What is the Council of Economic Advisers?
9. Why do economists sometimes offer conflicting advice to policymakers?

Problems and Applications

1. Describe some unusual language used in one of the other fields that you are studying. Why are these special terms useful?
2. One common assumption in economics is that the products of different firms in the same industry are indistinguishable. For each of the following industries, discuss whether this is a reasonable assumption.
 - a. steel
 - b. novels
 - c. wheat
 - d. fast food
3. Draw a circular-flow diagram. Identify the parts of the model that correspond to the flow of goods and services and the flow of dollars for each of the following activities.
 - a. Sam pays a storekeeper \$1 for a quart of milk.
 - b. Sally earns \$4.50 per hour working at a fast food restaurant.
 - c. Serena spends \$7 to see a movie.
 - d. Stuart earns \$10,000 from his 10 percent ownership of Acme Industrial.
4. Imagine a society that produces military goods and consumer goods, which we'll call "guns" and "butter."
 - a. Draw a production possibilities frontier for guns and butter. Explain why it most likely has a bowed-out shape.
 - b. Show a point that is impossible for the economy to achieve. Show a point that is feasible but inefficient.
 - c. Imagine that the society has two political parties, called the Hawks (who want a strong military) and the Doves (who want a smaller military). Show a point on your production possibilities frontier that the Hawks might choose and a point the Doves might choose.
 - d. Imagine that an aggressive neighboring country reduces the size of its military. As a result, both the Hawks and the Doves reduce their desired production of guns by the same amount. Which party would get the bigger "peace dividend," measured by the increase in butter production? Explain.
5. The first principle of economics discussed in Chapter 1 is that people face tradeoffs. Use a production possibilities frontier to illustrate society's tradeoff between a clean environment and high incomes. What do you suppose determines the shape and position of the frontier? Show what happens to the frontier if engineers develop an automobile engine with almost no emissions.
6. Classify the following topics as relating to microeconomics or macroeconomics.
 - a. a family's decision about how much income to save
 - b. the effect of government regulations on auto emissions
 - c. the impact of higher national saving on economic growth
 - d. a firm's decision about how many workers to hire
 - e. the relationship between the inflation rate and changes in the quantity of money
7. Classify each of the following statements as positive or normative. Explain.
 - a. Society faces a short-run tradeoff between inflation and unemployment.
 - b. A reduction in the rate of growth of money will reduce the rate of inflation.
 - c. The Federal Reserve should reduce the rate of growth of money.
 - d. Society ought to require welfare recipients to look for jobs.
 - e. Lower tax rates encourage more work and more saving.
8. Classify each of the statements in Table 2-2 as positive, normative, or ambiguous. Explain.
9. If you were president, would you be more interested in your economic advisers' positive views or their normative views? Why?
10. The *Economic Report of the President* contains statistical information about the economy as well as the Council of Economic Advisers' analysis of current policy issues. Find a recent copy of this annual report at your library and read a chapter about an issue that interests you. Summarize the economic problem at hand and describe the council's recommended policy.
11. Who is the current chairman of the Federal Reserve? Who is the current chair of the Council of Economic Advisers? Who is the current secretary of the treasury?
12. Look up one of the Web sites listed in Table 2-1. What recent economic trends or issues are addressed there?
13. Would you expect economists to disagree less about public policy as time goes on? Why or why not? Can their differences be completely eliminated? Why or why not?

APPENDIX

GRAPHING: A BRIEF REVIEW

Many of the concepts that economists study can be expressed with numbers—the price of bananas, the quantity of bananas sold, the cost of growing bananas, and so on. Often these economic variables are related to one another. When the price of bananas rises, people buy fewer bananas. One way of expressing the relationships among variables is with graphs.

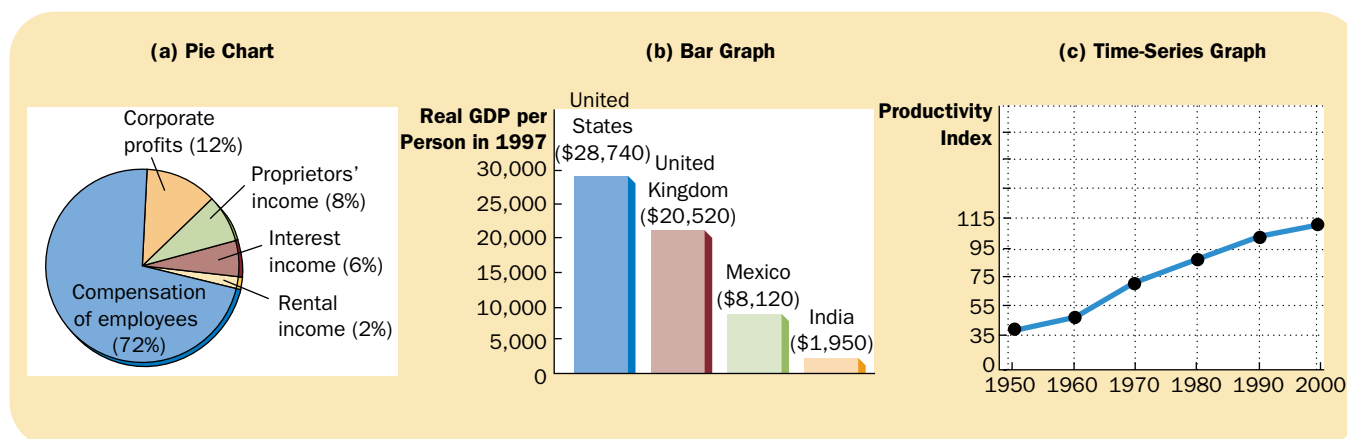
Graphs serve two purposes. First, when developing economic theories, graphs offer a way to visually express ideas that might be less clear if described with equations or words. Second, when analyzing economic data, graphs provide a way of finding how variables are in fact related in the world. Whether we are working with theory or with data, graphs provide a lens through which a recognizable forest emerges from a multitude of trees.

Numerical information can be expressed graphically in many ways, just as a thought can be expressed in words in many ways. A good writer chooses words that will make an argument clear, a description pleasing, or a scene dramatic. An effective economist chooses the type of graph that best suits the purpose at hand.

In this appendix we discuss how economists use graphs to study the mathematical relationships among variables. We also discuss some of the pitfalls that can arise in the use of graphical methods.

GRAPHS OF A SINGLE VARIABLE

Three common graphs are shown in Figure 2A-1. The *pie chart* in panel (a) shows how total income in the United States is divided among the sources of income, including compensation of employees, corporate profits, and so on. A slice of the pie represents each source's share of the total. The *bar graph* in panel (b) compares a measure of average income, called real GDP per person, for four countries. The height of each bar represents the average income in each country. The *time-series* graph in panel (c) traces the rising productivity in the U.S. business sector over time. The height of the line shows output per hour in each year. You have probably seen similar graphs presented in newspapers and magazines.



TYPES OF GRAPHS. The pie chart in panel (a) shows how U.S. national income is derived from various sources. The bar graph in panel (b) compares the average income in four countries. The time-series graph in panel (c) shows the growth in productivity of the U.S. business sector from 1950 to 2000.

Figure 2A-1

GRAPHS OF TWO VARIABLES: THE COORDINATE SYSTEM

Although the three graphs in Figure 2A-1 are useful in showing how a variable changes over time or across individuals, such graphs are limited in how much they can tell us. These graphs display information only on a single variable. Economists are often concerned with the relationships between variables. Thus, they need to be able to display two variables on a single graph. The *coordinate system* makes this possible.

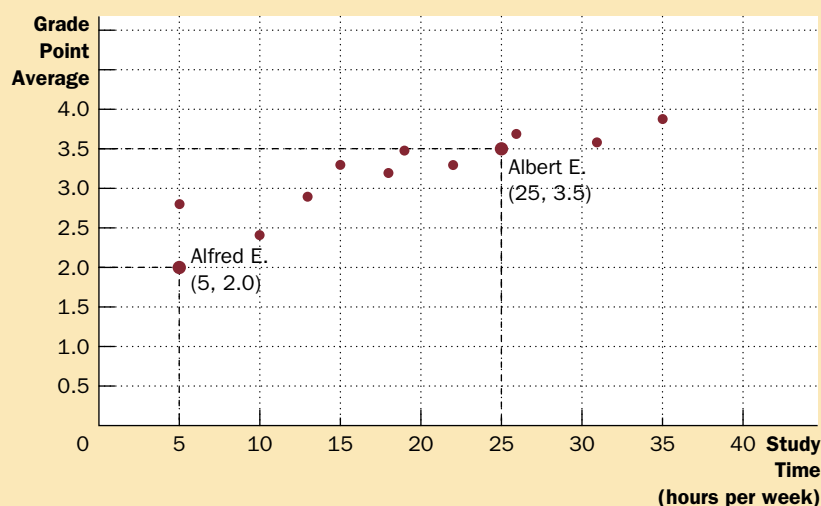
Suppose you want to examine the relationship between study time and grade point average. For each student in your class, you could record a pair of numbers: hours per week spent studying and grade point average. These numbers could then be placed in parentheses as an *ordered pair* and appear as a single point on the graph. Albert E., for instance, is represented by the ordered pair (25 hours/week, 3.5 GPA), while his “what-me-worry?” classmate Alfred E. is represented by the ordered pair (5 hours/week, 2.0 GPA).

We can graph these ordered pairs on a two-dimensional grid. The first number in each ordered pair, called the *x-coordinate*, tells us the horizontal location of the point. The second number, called the *y-coordinate*, tells us the vertical location of the point. The point with both an *x-coordinate* and a *y-coordinate* of zero is known as the *origin*. The two coordinates in the ordered pair tell us where the point is located in relation to the origin: *x* units to the right of the origin and *y* units above it.

Figure 2A-2 graphs grade point average against study time for Albert E., Alfred E., and their classmates. This type of graph is called a *scatterplot* because it plots scattered points. Looking at this graph, we immediately notice that points farther to the right (indicating more study time) also tend to be higher (indicating a better grade point average). Because study time and grade point average typically move in the same direction, we say that these two variables have a *positive*

Figure 2A-2

USING THE COORDINATE SYSTEM. Grade point average is measured on the vertical axis and study time on the horizontal axis. Albert E., Alfred E., and their classmates are represented by various points. We can see from the graph that students who study more tend to get higher grades.



correlation. By contrast, if we were to graph party time and grades, we would likely find that higher party time is associated with lower grades; because these variables typically move in opposite directions, we would call this a *negative correlation*. In either case, the coordinate system makes the correlation between the two variables easy to see.

CURVES IN THE COORDINATE SYSTEM

Students who study more do tend to get higher grades, but other factors also influence a student's grade. Previous preparation is an important factor, for instance, as are talent, attention from teachers, even eating a good breakfast. A scatterplot like Figure 2A-2 does not attempt to isolate the effect that study has on grades from the effects of other variables. Often, however, economists prefer looking at how one variable affects another holding everything else constant.

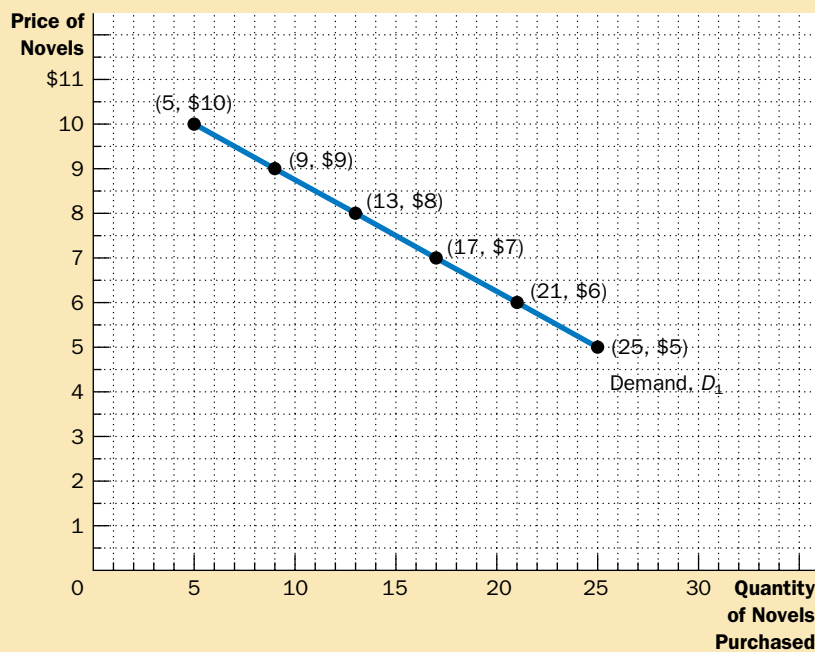
To see how this is done, let's consider one of the most important graphs in economics—the *demand curve*. The demand curve traces out the effect of a good's price on the quantity of the good consumers want to buy. Before showing a demand curve, however, consider Table 2A-1, which shows how the number of novels that Emma buys depends on her income and on the price of novels. When novels are cheap, Emma buys them in large quantities. As they become more expensive, she borrows books from the library instead of buying them or chooses to go to the movies instead of reading. Similarly, at any given price, Emma buys more novels when she has a higher income. That is, when her income increases, she spends part of the additional income on novels and part on other goods.

We now have three variables—the price of novels, income, and the number of novels purchased—which is more than we can represent in two dimensions. To

Table 2A-1

NOVELS PURCHASED BY EMMA. This table shows the number of novels Emma buys at various incomes and prices. For any given level of income, the data on price and quantity demanded can be graphed to produce Emma's demand curve for novels, as in Figure 2A-3.

PRICE	INCOME		
	\$20,000	\$30,000	\$40,000
\$10	2 novels	5 novels	8 novels
9	6	9	12
8	10	13	16
7	14	17	20
6	18	21	24
5	22	25	28
	Demand curve, D_3	Demand curve, D_1	Demand curve, D_2

**Figure 2A-3**

DEMAND CURVE. The line D_1 shows how Emma's purchases of novels depend on the price of novels when her income is held constant. Because the price and the quantity demanded are negatively related, the demand curve slopes downward.

put the information from Table 2A-1 in graphical form, we need to hold one of the three variables constant and trace out the relationship between the other two. Because the demand curve represents the relationship between price and quantity demanded, we hold Emma's income constant and show how the number of novels she buys varies with the price of novels.

Suppose that Emma's income is \$30,000 per year. If we place the number of novels Emma purchases on the x -axis and the price of novels on the y -axis, we can

graphically represent the middle column of Table 2A-1. When the points that represent these entries from the table—(5 novels, \$10), (9 novels, \$9), and so on—are connected, they form a line. This line, pictured in Figure 2A-3, is known as Emma's demand curve for novels; it tells us how many novels Emma purchases at any given price. The demand curve is downward sloping, indicating that a higher price reduces the quantity of novels demanded. Because the quantity of novels demanded and the price move in opposite directions, we say that the two variables are *negatively related*. (Conversely, when two variables move in the same direction, the curve relating them is upward sloping, and we say the variables are *positively related*.)

Now suppose that Emma's income rises to \$40,000 per year. At any given price, Emma will purchase more novels than she did at her previous level of income. Just as earlier we drew Emma's demand curve for novels using the entries from the middle column of Table 2A-1, we now draw a new demand curve using the entries from the right-hand column of the table. This new demand curve (curve D_2) is pictured alongside the old one (curve D_1) in Figure 2A-4; the new curve is a similar line drawn farther to the right. We therefore say that Emma's demand curve for novels *shifts* to the right when her income increases. Likewise, if Emma's income were to fall to \$20,000 per year, she would buy fewer novels at any given price and her demand curve would shift to the left (to curve D_3).

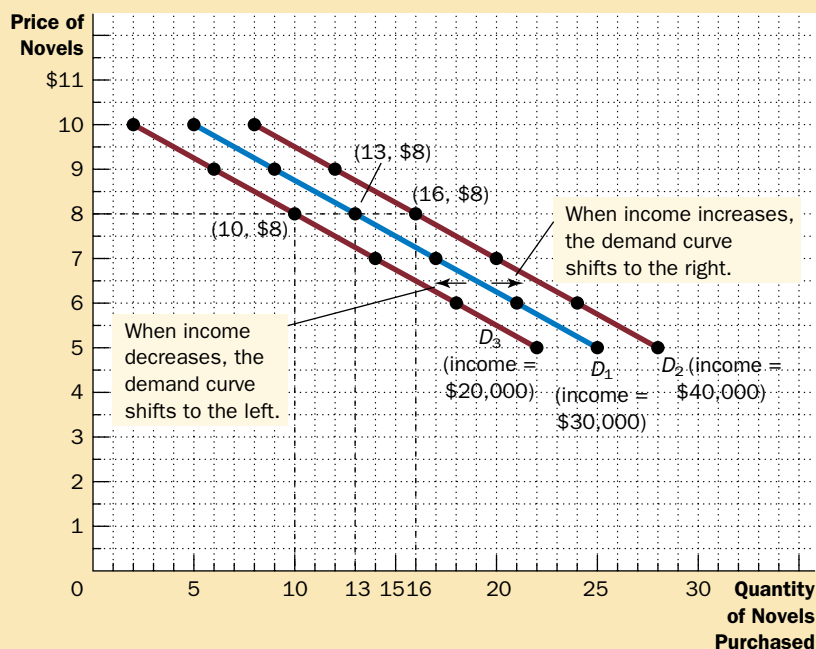
In economics, it is important to distinguish between *movements along a curve* and *shifts of a curve*. As we can see from Figure 2A-3, if Emma earns \$30,000 per year and novels cost \$8 apiece, she will purchase 13 novels per year. If the price of novels falls to \$7, Emma will increase her purchases of novels to 17 per year. The demand curve, however, stays fixed in the same place. Emma still buys the same

Figure 2A-4

SHIFTING DEMAND CURVES.

The location of Emma's demand curve for novels depends on how much income she earns. The more she earns, the more novels she will purchase at any given price, and the farther to the right her demand curve will lie.

Curve D_1 represents Emma's original demand curve when her income is \$30,000 per year. If her income rises to \$40,000 per year, her demand curve shifts to D_2 . If her income falls to \$20,000 per year, her demand curve shifts to D_3 .



number of novels at *each price*, but as the price falls she moves along her demand curve from left to right. By contrast, if the price of novels remains fixed at \$8 but her income rises to \$40,000, Emma increases her purchases of novels from 13 to 16 per year. Because Emma buys more novels *at each price*, her demand curve shifts out, as shown in Figure 2A-4.

There is a simple way to tell when it is necessary to shift a curve. When a variable that is not named on either axis changes, the curve shifts. Income is on neither the x -axis nor the y -axis of the graph, so when Emma's income changes, her demand curve must shift. Any change that affects Emma's purchasing habits besides a change in the price of novels will result in a shift in her demand curve. If, for instance, the public library closes and Emma must buy all the books she wants to read, she will demand more novels at each price, and her demand curve will shift to the right. Or, if the price of movies falls and Emma spends more time at the movies and less time reading, she will demand fewer novels at each price, and her demand curve will shift to the left. By contrast, when a variable on an axis of the graph changes, the curve does not shift. We read the change as a movement along the curve.

SLOPE

One question we might want to ask about Emma is how much her purchasing habits respond to price. Look at the demand curve pictured in Figure 2A-5. If this curve is very steep, Emma purchases nearly the same number of novels regardless

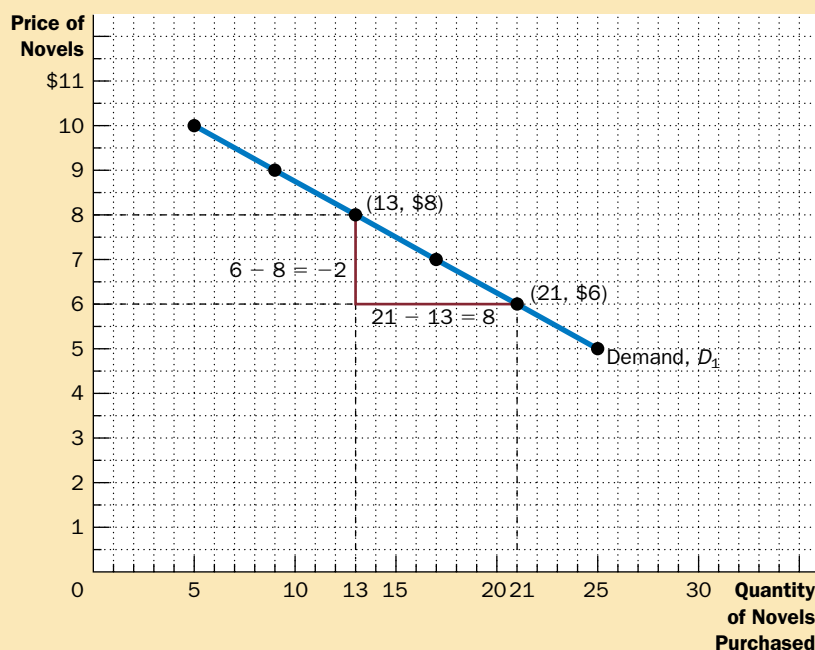


Figure 2A-5

CALCULATING THE SLOPE OF A LINE. To calculate the slope of the demand curve, we can look at the changes in the x - and y -coordinates as we move from the point (21 novels, \$6) to the point (13 novels, \$8). The slope of the line is the ratio of the change in the y -coordinate (-2) to the change in the x -coordinate ($+8$), which equals $-1/4$.

of whether they are cheap or expensive. If this curve is much flatter, Emma purchases many fewer novels when the price rises. To answer questions about how much one variable responds to changes in another variable, we can use the concept of *slope*.

The slope of a line is the ratio of the vertical distance covered to the horizontal distance covered as we move along the line. This definition is usually written out in mathematical symbols as follows:

$$\text{slope} = \frac{\Delta y}{\Delta x},$$

where the Greek letter Δ (delta) stands for the change in a variable. In other words, the slope of a line is equal to the “rise” (change in y) divided by the “run” (change in x). The slope will be a small positive number for a fairly flat upward-sloping line, a large positive number for a steep upward-sloping line, and a negative number for a downward-sloping line. A horizontal line has a slope of zero because in this case the y -variable never changes; a vertical line is defined to have an infinite slope because the y -variable can take any value without the x -variable changing at all.

What is the slope of Emma’s demand curve for novels? First of all, because the curve slopes down, we know the slope will be negative. To calculate a numerical value for the slope, we must choose two points on the line. With Emma’s income at \$30,000, she will purchase 21 novels at a price of \$6 or 13 novels at a price of \$8. When we apply the slope formula, we are concerned with the change between these two points; in other words, we are concerned with the difference between them, which lets us know that we will have to subtract one set of values from the other, as follows:

$$\text{slope} = \frac{\Delta y}{\Delta x} = \frac{\text{first } y\text{-coordinate} - \text{second } y\text{-coordinate}}{\text{first } x\text{-coordinate} - \text{second } x\text{-coordinate}} = \frac{6 - 8}{21 - 13} = \frac{-2}{8} = \frac{-1}{4}.$$

Figure 2A-5 shows graphically how this calculation works. Try computing the slope of Emma’s demand curve using two different points. You should get exactly the same result, $-1/4$. One of the properties of a straight line is that it has the same slope everywhere. This is not true of other types of curves, which are steeper in some places than in others.

The slope of Emma’s demand curve tells us something about how responsive her purchases are to changes in the price. A small slope (a number close to zero) means that Emma’s demand curve is relatively flat; in this case, she adjusts the number of novels she buys substantially in response to a price change. A larger slope (a number farther from zero) means that Emma’s demand curve is relatively steep; in this case, she adjusts the number of novels she buys only slightly in response to a price change.

CAUSE AND EFFECT

Economists often use graphs to advance an argument about how the economy works. In other words, they use graphs to argue about how one set of events *causes* another set of events. With a graph like the demand curve, there is no doubt about cause and effect. Because we are varying price and holding all other

variables constant, we know that changes in the price of novels cause changes in the quantity Emma demands. Remember, however, that our demand curve came from a hypothetical example. When graphing data from the real world, it is often more difficult to establish how one variable affects another.

The first problem is that it is difficult to hold everything else constant when measuring how one variable affects another. If we are not able to hold variables constant, we might decide that one variable on our graph is causing changes in the other variable when actually those changes are caused by a third *omitted* variable not pictured on the graph. Even if we have identified the correct two variables to look at, we might run into a second problem—*reverse causality*. In other words, we might decide that A causes B when in fact B causes A. The omitted-variable and reverse-causality traps require us to proceed with caution when using graphs to draw conclusions about causes and effects.

Omitted Variables To see how omitting a variable can lead to a deceptive graph, let's consider an example. Imagine that the government, spurred by public concern about the large number of deaths from cancer, commissions an exhaustive study from Big Brother Statistical Services, Inc. Big Brother examines many of the items found in people's homes to see which of them are associated with the risk of cancer. Big Brother reports a strong relationship between two variables: the number of cigarette lighters that a household owns and the probability that someone in the household will develop cancer. Figure 2A-6 shows this relationship.

What should we make of this result? Big Brother advises a quick policy response. It recommends that the government discourage the ownership of cigarette lighters by taxing their sale. It also recommends that the government require warning labels: "Big Brother has determined that this lighter is dangerous to your health."

In judging the validity of Big Brother's analysis, one question is paramount: Has Big Brother held constant every relevant variable except the one under consideration? If the answer is no, the results are suspect. An easy explanation for Figure 2A-6 is that people who own more cigarette lighters are more likely to smoke cigarettes and that cigarettes, not lighters, cause cancer. If Figure 2A-6 does not

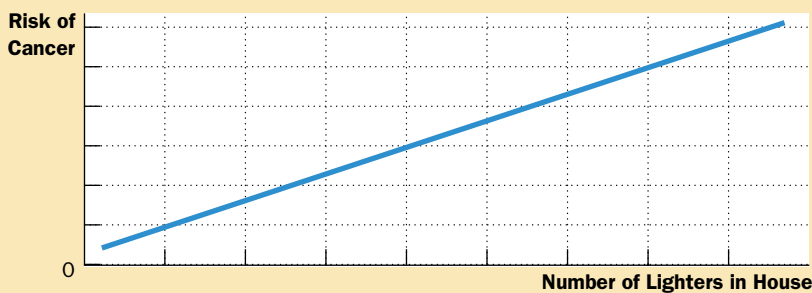


Figure 2A-6

GRAPH WITH AN OMITTED VARIABLE. The upward-sloping curve shows that members of households with more cigarette lighters are more likely to develop cancer. Yet we should not conclude that ownership of lighters causes cancer because the graph does not take into account the number of cigarettes smoked.

hold constant the amount of smoking, it does not tell us the true effect of owning a cigarette lighter.

This story illustrates an important principle: When you see a graph being used to support an argument about cause and effect, it is important to ask whether the movements of an omitted variable could explain the results you see.

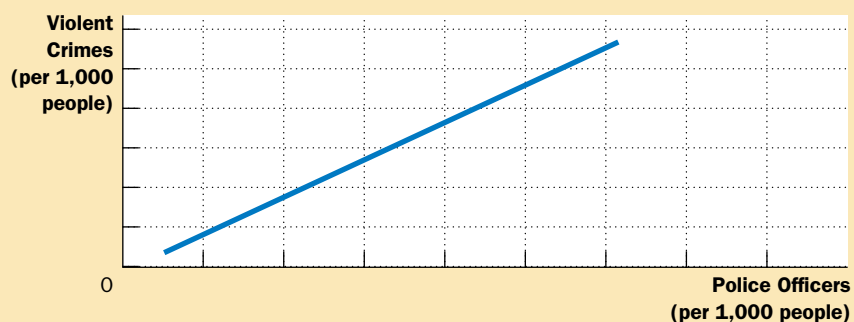
Reverse Causality Economists can also make mistakes about causality by misreading its direction. To see how this is possible, suppose the Association of American Anarchists commissions a study of crime in America and arrives at Figure 2A-7, which plots the number of violent crimes per thousand people in major cities against the number of police officers per thousand people. The anarchists note the curve's upward slope and argue that because police increase rather than decrease the amount of urban violence, law enforcement should be abolished.

If we could run a controlled experiment, we would avoid the danger of reverse causality. To run an experiment, we would set the number of police officers in different cities randomly and then examine the correlation between police and crime. Figure 2A-7, however, is not based on such an experiment. We simply observe that more dangerous cities have more police officers. The explanation for this may be that more dangerous cities hire more police. In other words, rather than police causing crime, crime may cause police. Nothing in the graph itself allows us to establish the direction of causality.

It might seem that an easy way to determine the direction of causality is to examine which variable moves first. If we see crime increase and then the police force expand, we reach one conclusion. If we see the police force expand and then crime increase, we reach the other. Yet there is also a flaw with this approach: Often people change their behavior not in response to a change in their present conditions but in response to a change in their *expectations* of future conditions. A city that expects a major crime wave in the future, for instance, might well hire more police now. This problem is even easier to see in the case of babies and minivans. Couples often buy a minivan in anticipation of the birth of a child. The

Figure 2A-7

GRAPH SUGGESTING REVERSE CAUSALITY. The upward-sloping curve shows that cities with a higher concentration of police are more dangerous. Yet the graph does not tell us whether police cause crime or crime-plagued cities hire more police.



minivan comes before the baby, but we wouldn't want to conclude that the sale of minivans causes the population to grow!

There is no complete set of rules that says when it is appropriate to draw causal conclusions from graphs. Yet just keeping in mind that cigarette lighters don't cause cancer (omitted variable) and minivans don't cause larger families (reverse causality) will keep you from falling for many faulty economic arguments.

3



INTERDEPENDENCE AND THE GAINS FROM TRADE

Consider your typical day. You wake up in the morning, and you pour yourself juice from oranges grown in Florida and coffee from beans grown in Brazil. Over breakfast, you watch a news program broadcast from New York on your television made in Japan. You get dressed in clothes made of cotton grown in Georgia and sewn in factories in Thailand. You drive to class in a car made of parts manufactured in more than a dozen countries around the world. Then you open up your economics textbook written by an author living in Massachusetts, published by a company located in Texas, and printed on paper made from trees grown in Oregon.

Every day you rely on many people from around the world, most of whom you do not know, to provide you with the goods and services that you enjoy. Such interdependence is possible because people trade with one another. Those people who provide you with goods and services are not acting out of generosity or concern for your welfare. Nor is some government agency directing them to make what you

**IN THIS CHAPTER
YOU WILL . . .**

***Consider how
everyone can benefit
when people trade
with one another***

***Learn the meaning of
absolute advantage
and comparative
advantage***

***See how comparative
advantage explains
the gains from trade***

***Apply the theory of
comparative
advantage to
everyday life and
national policy***

want and to give it to you. Instead, people provide you and other consumers with the goods and services they produce because they get something in return.

In subsequent chapters we will examine how our economy coordinates the activities of millions of people with varying tastes and abilities. As a starting point for this analysis, here we consider the reasons for economic interdependence. One of the *Ten Principles of Economics* highlighted in Chapter 1 is that trade can make everyone better off. This principle explains why people trade with their neighbors and why nations trade with other nations. In this chapter we examine this principle more closely. What exactly do people gain when they trade with one another? Why do people choose to become interdependent?



A PARABLE FOR THE MODERN ECONOMY

To understand why people choose to depend on others for goods and services and how this choice improves their lives, let's look at a simple economy. Imagine that there are two goods in the world—meat and potatoes. And there are two people in the world—a cattle rancher and a potato farmer—each of whom would like to eat both meat and potatoes.

The gains from trade are most obvious if the rancher can produce only meat and the farmer can produce only potatoes. In one scenario, the rancher and the farmer could choose to have nothing to do with each other. But after several months of eating beef roasted, boiled, broiled, and grilled, the rancher might decide that self-sufficiency is not all it's cracked up to be. The farmer, who has been eating potatoes mashed, fried, baked, and scalloped, would likely agree. It is easy to see that trade would allow them to enjoy greater variety: Each could then have a hamburger with french fries.

Although this scene illustrates most simply how everyone can benefit from trade, the gains would be similar if the rancher and the farmer were each capable of producing the other good, but only at great cost. Suppose, for example, that the potato farmer is able to raise cattle and produce meat, but that he is not very good at it. Similarly, suppose that the cattle rancher is able to grow potatoes, but that her land is not very well suited for it. In this case, it is easy to see that the farmer and the rancher can each benefit by specializing in what he or she does best and then trading with the other.

The gains from trade are less obvious, however, when one person is better at producing *every* good. For example, suppose that the rancher is better at raising cattle *and* better at growing potatoes than the farmer. In this case, should the rancher or farmer choose to remain self-sufficient? Or is there still reason for them to trade with each other? To answer this question, we need to look more closely at the factors that affect such a decision.

PRODUCTION POSSIBILITIES

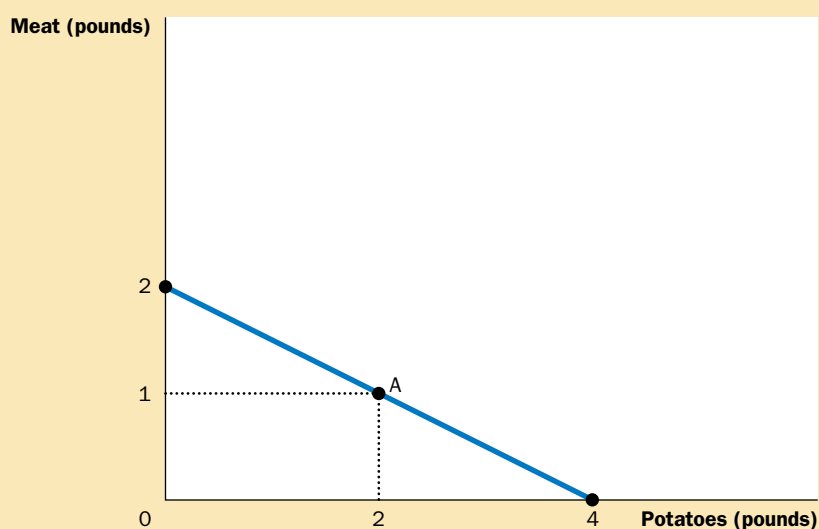
Suppose that the farmer and the rancher each work 40 hours a week and can devote this time to growing potatoes, raising cattle, or a combination of the two. Table 3-1 shows the amount of time each person requires to produce 1 pound of

Table 3-1

THE PRODUCTION OPPORTUNITIES OF THE FARMER AND THE RANCHER

	HOURS NEEDED TO MAKE 1 POUND OF:		AMOUNT PRODUCED IN 40 HOURS	
	MEAT	POTATOES	MEAT	POTATOES
FARMER	20 hours/lb	10 hours/lb	2 lbs	4 lbs
RANCHER	1 hour/lb	8 hours/lb	40 lbs	5 lbs

(a) The Farmer's Production Possibilities Frontier



(b) The Rancher's Production Possibilities Frontier

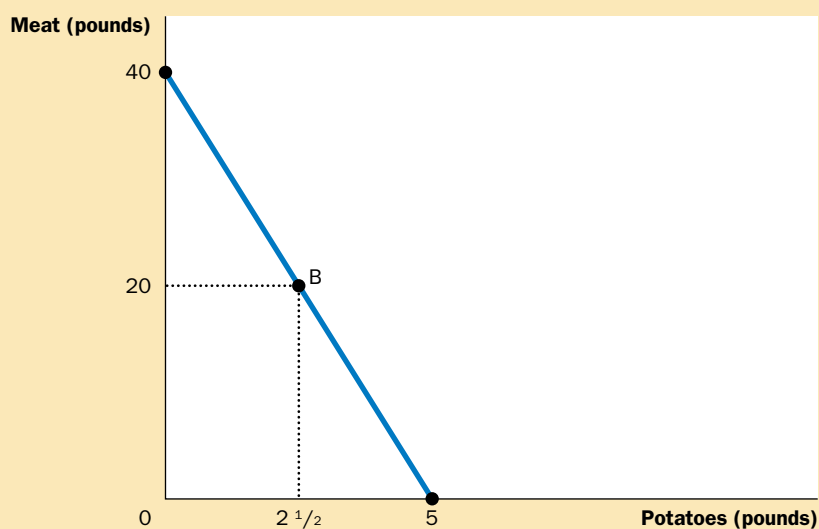


Figure 3-1

THE PRODUCTION POSSIBILITIES FRONTIER. Panel (a) shows the combinations of meat and potatoes that the farmer can produce. Panel (b) shows the combinations of meat and potatoes that the rancher can produce. Both production possibilities frontiers are derived from Table 3-1 and the assumption that the farmer and rancher each work 40 hours per week.

each good. The farmer can produce a pound of potatoes in 10 hours and a pound of meat in 20 hours. The rancher, who is more productive in both activities, can produce a pound of potatoes in 8 hours and a pound of meat in 1 hour.

Panel (a) of Figure 3-1 illustrates the amounts of meat and potatoes that the farmer can produce. If the farmer devotes all 40 hours of his time to potatoes, he produces 4 pounds of potatoes and no meat. If he devotes all his time to meat, he produces 2 pounds of meat and no potatoes. If the farmer divides his time equally between the two activities, spending 20 hours on each, he produces 2 pounds of potatoes and 1 pound of meat. The figure shows these three possible outcomes and all others in between.



This graph is the farmer's production possibilities frontier. As we discussed in Chapter 2, a production possibilities frontier shows the various mixes of output that an economy can produce. It illustrates one of the *Ten Principles of Economics* in Chapter 1: People face tradeoffs. Here the farmer faces a tradeoff between producing meat and producing potatoes. You may recall that the production possibilities frontier in Chapter 2 was drawn bowed out; in this case, the tradeoff between the two goods depends on the amounts being produced. Here, however, the farmer's technology for producing meat and potatoes (as summarized in Table 3-1) allows him to switch between one good and the other at a constant rate. In this case, the production possibilities frontier is a straight line.

Panel (b) of Figure 3-1 shows the production possibilities frontier for the rancher. If the rancher devotes all 40 hours of her time to potatoes, she produces 5 pounds of potatoes and no meat. If she devotes all her time to meat, she produces 40 pounds of meat and no potatoes. If the rancher divides her time equally, spending 20 hours on each activity, she produces 2 1/2 pounds of potatoes and 20 pounds of meat. Once again, the production possibilities frontier shows all the possible outcomes.

If the farmer and rancher choose to be self-sufficient, rather than trade with each other, then each consumes exactly what he or she produces. In this case, the production possibilities frontier is also the consumption possibilities frontier. That is, without trade, Figure 3-1 shows the possible combinations of meat and potatoes that the farmer and rancher can each consume.

Although these production possibilities frontiers are useful in showing the tradeoffs that the farmer and rancher face, they do not tell us what the farmer and rancher will actually choose to do. To determine their choices, we need to know the tastes of the farmer and the rancher. Let's suppose they choose the combinations identified by points A and B in Figure 3-1: The farmer produces and consumes 2 pounds of potatoes and 1 pound of meat, while the rancher produces and consumes 2 1/2 pounds of potatoes and 20 pounds of meat.

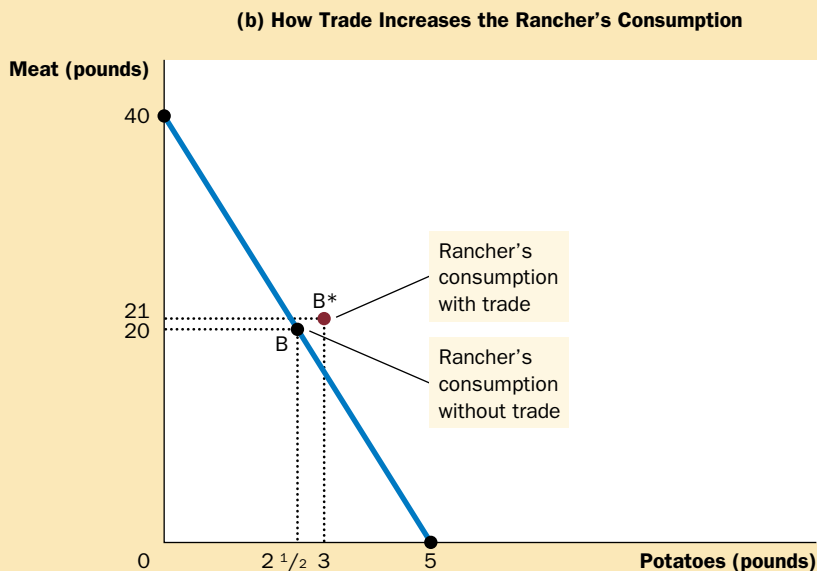
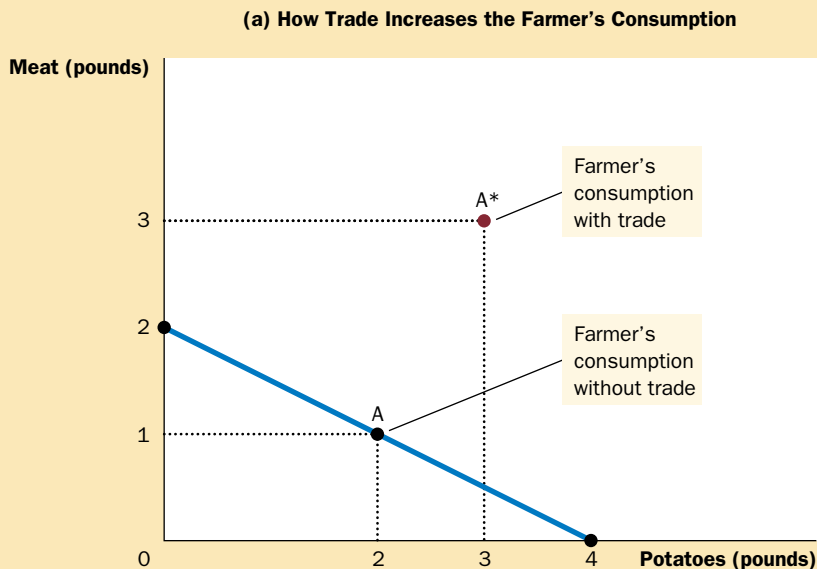
SPECIALIZATION AND TRADE

After several years of eating combination B, the rancher gets an idea and goes to talk to the farmer:

RANCHER: Farmer, my friend, have I got a deal for you! I know how to improve life for both of us. I think you should stop producing meat altogether and devote all your time to growing potatoes. According to my calculations, if you work 40 hours a week growing potatoes, you'll

Figure 3-2

HOW TRADE EXPANDS THE SET OF CONSUMPTION OPPORTUNITIES. The proposed trade between the farmer and the rancher offers each of them a combination of meat and potatoes that would be impossible in the absence of trade. In panel (a), the farmer gets to consume at point A* rather than point A. In panel (b), the rancher gets to consume at point B* rather than point B. Trade allows each to consume more meat and more potatoes.



produce 4 pounds of potatoes. If you give me 1 of those 4 pounds, I'll give you 3 pounds of meat in return. In the end, you'll get to eat 3 pounds of potatoes and 3 pounds of meat every week, instead of the 2 pounds of potatoes and 1 pound of meat you now get. If you go along with my plan, you'll have more of *both* foods. [To illustrate her point, the rancher shows the farmer panel (a) of Figure 3-2.]

FARMER: (*sounding skeptical*) That seems like a good deal for me. But I don't understand why you are offering it. If the deal is so good for me, it can't be good for you too.

	THE OUTCOME WITHOUT TRADE:		THE OUTCOME WITH TRADE:			THE GAINS FROM TRADE:		
	WHAT THEY PRODUCE AND CONSUME		WHAT THEY PRODUCE	WHAT THEY TRADE	WHAT THEY CONSUME	THE INCREASE IN CONSUMPTION		
FARMER	1 lb meat 2 lbs potatoes	}point A	0 lbs meat 4 lbs potatoes	Gets 3 lbs meat for 1 lb potatoes	3 lbs meat 3 lbs potatoes	}point A*	2 lbs meat 1 lb potatoes	}A* - A
RANCHER	20 lbs meat 2 1/2 lbs potatoes	}point B	24 lbs meat 2 lbs potatoes	Gives 3 lbs meat for 1 lb potatoes	21 lbs meat 3 lbs potatoes	}point B*	1 lb meat 1/2 lb potatoes	}B* - B

Table 3-2

THE GAINS FROM TRADE: A SUMMARY

RANCHER: Oh, but it is! If I spend 24 hours a week raising cattle and 16 hours growing potatoes, I'll produce 24 pounds of meat and 2 pounds of potatoes. After I give you 3 pounds of meat in exchange for 1 pound of potatoes, I'll have 21 pounds of meat and 3 pounds of potatoes. In the end, I will also get more of both foods than I have now. [She points out panel (b) of Figure 3-2.]

FARMER: I don't know. . . . This sounds too good to be true.

RANCHER: It's really not as complicated as it seems at first. Here—I have summarized my proposal for you in a simple table. [The rancher hands the farmer a copy of Table 3-2.]

FARMER: *(after pausing to study the table)* These calculations seem correct, but I am puzzled. How can this deal make us both better off?

RANCHER: We can both benefit because trade allows each of us to specialize in doing what we do best. You will spend more time growing potatoes and less time raising cattle. I will spend more time raising cattle and less time growing potatoes. As a result of specialization and trade, each of us can consume both more meat and more potatoes without working any more hours.

QUICK QUIZ: Draw an example of a production possibilities frontier for Robinson Crusoe, a shipwrecked sailor who spends his time gathering coconuts and catching fish. Does this frontier limit Crusoe's consumption of coconuts and fish if he lives by himself? Does he face the same limits if he can trade with natives on the island?

THE PRINCIPLE OF COMPARATIVE ADVANTAGE

The rancher's explanation of the gains from trade, though correct, poses a puzzle: If the rancher is better at both raising cattle and growing potatoes, how can the farmer ever specialize in doing what he does best? The farmer doesn't seem to do

anything best. To solve this puzzle, we need to look at the principle of *comparative advantage*.

As a first step in developing this principle, consider the following question: In our example, who can produce potatoes at lower cost—the farmer or the rancher? There are two possible answers, and in these two answers lie both the solution to our puzzle and the key to understanding the gains from trade.

ABSOLUTE ADVANTAGE

One way to answer the question about the cost of producing potatoes is to compare the inputs required by the two producers. The rancher needs only 8 hours to produce a pound of potatoes, whereas the farmer needs 10 hours. Based on this information, one might conclude that the rancher has the lower cost of producing potatoes.

Economists use the term **absolute advantage** when comparing the productivity of one person, firm, or nation to that of another. The producer that requires a smaller quantity of inputs to produce a good is said to have an absolute advantage in producing that good. In our example, the rancher has an absolute advantage both in producing potatoes and in producing meat, because she requires less time than the farmer to produce a unit of either good.

absolute advantage

the comparison among producers of a good according to their productivity

OPPORTUNITY COST AND COMPARATIVE ADVANTAGE

There is another way to look at the cost of producing potatoes. Rather than comparing inputs required, we can compare the opportunity costs. Recall from Chapter 1 that the **opportunity cost** of some item is what we give up to get that item. In our example, we assumed that the farmer and the rancher each spend 40 hours a week working. Time spent producing potatoes, therefore, takes away from time available for producing meat. As the rancher and farmer change their allocations of time between producing the two goods, they move along their production possibility frontiers; in a sense, they are using one good to produce the other. The opportunity cost measures the tradeoff that each of them faces.

Let's first consider the rancher's opportunity cost. Producing 1 pound of potatoes takes her 8 hours of work. When the rancher spends that 8 hours producing potatoes, she spends 8 hours less producing meat. Because the rancher needs only 1 hour to produce 1 pound of meat, 8 hours of work would yield 8 pounds of meat. Hence, the rancher's opportunity cost of 1 pound of potatoes is 8 pounds of meat.

Now consider the farmer's opportunity cost. Producing 1 pound of potatoes takes him 10 hours. Because he needs 20 hours to produce 1 pound of meat, 10 hours would yield 1/2 pound of meat. Hence, the farmer's opportunity cost of 1 pound of potatoes is 1/2 pound of meat.

Table 3-3 shows the opportunity cost of meat and potatoes for the two producers. Notice that the opportunity cost of meat is the inverse of the opportunity cost of potatoes. Because 1 pound of potatoes costs the rancher 8 pounds of meat, 1 pound of meat costs the rancher 1/8 pound of potatoes. Similarly, because 1 pound of potatoes costs the farmer 1/2 pound of meat, 1 pound of meat costs the farmer 2 pounds of potatoes.

Economists use the term **comparative advantage** when describing the opportunity cost of two producers. The producer who has the smaller opportunity cost

opportunity cost

whatever must be given up to obtain some item

comparative advantage

the comparison among producers of a good according to their opportunity cost

Table 3-3THE OPPORTUNITY COST OF
MEAT AND POTATOES

	OPPORTUNITY COST OF:	
	1 POUND OF MEAT	1 POUND OF POTATOES
FARMER	2 lbs potatoes	1/2 lb meat
RANCHER	1/8 lb potatoes	8 lbs meat

of producing a good—that is, who has to give up less of other goods to produce it—is said to have a comparative advantage in producing that good. In our example, the farmer has a lower opportunity cost of producing potatoes than the rancher (1/2 pound versus 8 pounds of meat). The rancher has a lower opportunity cost of producing meat than the farmer (1/8 pound versus 2 pounds of potatoes). Thus, the farmer has a comparative advantage in growing potatoes, and the rancher has a comparative advantage in producing meat.

Notice that it would be impossible for the same person to have a comparative advantage in both goods. Because the opportunity cost of one good is the inverse of the opportunity cost of the other, if a person's opportunity cost of one good is relatively high, his opportunity cost of the other good must be relatively low. Comparative advantage reflects the relative opportunity cost. Unless two people have exactly the same opportunity cost, one person will have a comparative advantage in one good, and the other person will have a comparative advantage in the other good.

COMPARATIVE ADVANTAGE AND TRADE

Differences in opportunity cost and comparative advantage create the gains from trade. When each person specializes in producing the good for which he or she has a comparative advantage, total production in the economy rises, and this increase in the size of the economic pie can be used to make everyone better off. In other words, as long as two people have different opportunity costs, each can benefit from trade by obtaining a good at a price lower than his or her opportunity cost of that good.

Consider the proposed deal from the viewpoint of the farmer. The farmer gets 3 pounds of meat in exchange for 1 pound of potatoes. In other words, the farmer buys each pound of meat for a price of 1/3 pound of potatoes. This price of meat is lower than his opportunity cost for 1 pound of meat, which is 2 pounds of potatoes. Thus, the farmer benefits from the deal because he gets to buy meat at a good price.

Now consider the deal from the rancher's viewpoint. The rancher buys 1 pound of potatoes for a price of 3 pounds of meat. This price of potatoes is lower than her opportunity cost of 1 pound of potatoes, which is 8 pounds of meat. Thus, the rancher benefits because she gets to buy potatoes at a good price.

These benefits arise because each person concentrates on the activity for which he or she has the lower opportunity cost: The farmer spends more time growing potatoes, and the rancher spends more time producing meat. As a result, the total production of potatoes and the total production of meat both rise, and the farmer

FYI*The Legacy of Adam Smith and David Ricardo*

Economists have long understood the principle of comparative advantage. Here is how the great economist Adam Smith put the argument:

It is a maxim of every prudent master of a family, never to attempt to make at home what it will cost him more to make than to buy. The tailor does not attempt to make his own

shoes, but buys them of the shoemaker. The shoemaker does not attempt to make his own clothes but employs a tailor. The farmer attempts to make neither the one nor the other, but employs those different artificers. All of them find it for their interest to employ their whole industry in a way in which they have some advantage over their neighbors, and to purchase with a part of its produce, or what is the same thing, with the price of part of it, whatever else they have occasion for.

This quotation is from Smith's 1776 book, *An Inquiry into the Nature and Causes of the Wealth of Nations*, which was

a landmark in the analysis of trade and economic interdependence.

Smith's book inspired David Ricardo, a millionaire stockbroker, to become an economist. In his 1817 book, *Principles of Political Economy and Taxation*, Ricardo developed the principle of comparative advantage as we know it today. His defense of free trade was not a mere academic exercise. Ricardo put his economic beliefs to work as a member of the British Parliament, where he opposed the Corn Laws, which restricted the import of grain.



DAVID RICARDO

The conclusions of Adam Smith and David Ricardo on the gains from trade have held up well over time. Although economists often disagree on questions of policy, they are united in their support of free trade. Moreover, the central argument for free trade has not changed much in the past two centuries. Even though the field of economics has broadened its scope and refined its theories since the time of Smith and Ricardo, economists' opposition to trade restrictions is still based largely on the principle of comparative advantage.

and rancher share the benefits of this increased production. The moral of the story of the farmer and the rancher should now be clear: *Trade can benefit everyone in society because it allows people to specialize in activities in which they have a comparative advantage.*

QUICK QUIZ: Robinson Crusoe can gather 10 coconuts or catch 1 fish per hour. His friend Friday can gather 30 coconuts or catch 2 fish per hour. What is Crusoe's opportunity cost of catching one fish? What is Friday's? Who has an absolute advantage in catching fish? Who has a comparative advantage in catching fish?

APPLICATIONS OF COMPARATIVE ADVANTAGE

The principle of comparative advantage explains interdependence and the gains from trade. Because interdependence is so prevalent in the modern world, the principle of comparative advantage has many applications. Here are two examples, one fanciful and one of great practical importance.

SHOULD TIGER WOODS MOW HIS OWN LAWN?

Tiger Woods spends a lot of time walking around on grass. One of the most talented golfers of all time, he can hit a drive and sink a putt in a way that most casual golfers only dream of doing. Most likely, he is talented at other activities too. For example, let's imagine that Woods can mow his lawn faster than anyone else. But just because he *can* mow his lawn fast, does this mean he *should*?

To answer this question, we can use the concepts of opportunity cost and comparative advantage. Let's say that Woods can mow his lawn in 2 hours. In that same 2 hours, he could film a television commercial for Nike and earn \$10,000. By contrast, Forrest Gump, the boy next door, can mow Woods's lawn in 4 hours. In that same 4 hours, he could work at McDonald's and earn \$20.

In this example, Woods's opportunity cost of mowing the lawn is \$10,000 and Forrest's opportunity cost is \$20. Woods has an absolute advantage in mowing lawns because he can do the work in less time. Yet Forrest has a comparative advantage in mowing lawns because he has the lower opportunity cost.

IN THE NEWS

Who has a Comparative Advantage in Producing Lamb?



A COMMON BARRIER TO FREE TRADE among countries is tariffs, which are taxes on the import of goods from abroad. In the following opinion column, economist Douglas Irwin discusses a recent example of their use.

Lamb Tariffs Fleece U.S. Consumers

BY DOUGLAS A. IRWIN

President Clinton dealt a serious blow to free trade last Wednesday, when he announced that the U.S. would impose stiff import tariffs on lamb from Australia and New Zealand. His decision undercuts

American leadership and makes a mockery of the administration's claims that it favors free and fair trade.

U.S. sheep producers have long been dependent on government. For more than half a century, until Congress enacted farm-policy reforms in 1995, they received subsidies for wool. Having lost that handout, saddled with high costs and inefficiencies, and facing domestic competition from chicken, beef, and pork, sheep producers sought to stop foreign competition by filing for import relief.

Almost all U.S. lamb imports come from Australia and New Zealand, major agricultural producers with a crushing comparative advantage. New Zealand has fewer than four million people but as many as 60 million sheep (compared with about seven million sheep in the U.S.). New Zealand's farmers have invested substantial resources in new technology and effective marketing, making them among the most efficient producers in the world. New Zealand also eliminated domestic agricultural

subsidies in the free-market reforms of the 1950s, and is a free-trading country, on track to eliminate all import tariffs by 2006.

Rather than emulate this example, the American Sheep Industry Association, among others, filed an "escape clause" petition under the Trade Act of 1974, which allows temporary "breathing space" protection to import-competing industries. Under the escape-clause provision, a petitioning industry is required to present an adjustment plan to ensure that it undertakes steps to become competitive in the future. The tariff protection is usually limited and scheduled to be phased out.

The U.S. International Trade Commission determines whether imports are a cause of "serious injury" to the domestic industry and, if so, proposes a remedy, which the president has full discretion to adopt, change or reject. In February, the ITC did not find that the domestic industry had suffered "serious injury," but rather adopted the weaker ruling that imports were "a substantial

The gains from trade in this example are tremendous. Rather than mowing his own lawn, Woods should make the commercial and hire Forrest to mow the lawn. As long as Woods pays Forrest more than \$20 and less than \$10,000, both of them are better off.

SHOULD THE UNITED STATES TRADE WITH OTHER COUNTRIES?

Just as individuals can benefit from specialization and trade with one another, as the farmer and rancher did, so can populations of people in different countries. Many of the goods that Americans enjoy are produced abroad, and many of the goods produced in the United States are sold abroad. Goods produced abroad and sold domestically are called **imports**. Goods produced domestically and sold abroad are called **exports**.

imports

goods produced abroad and sold domestically

exports

goods produced domestically and sold abroad

cause of threat of serious injury." The ITC did not propose to roll back imports, only to impose a 20% tariff (declining over four years) on imports above last year's levels.

The administration at first appeared to be considering less restrictive measures. Australia and New Zealand even offered financial assistance to the U.S. producers, and the administration delayed any announcement and appeared to be working toward a compromise. But these hopes were completely dashed with the shocking final decision, in which the administration capitulated to the demands of the sheep industry and its advocates in Congress.

The congressional charge was led by Sen. Max Baucus (D., Mont.), a member of the Agriculture Committee whose sister, a sheep producer, had appeared before the ITC to press for higher tariffs. The administration opted for . . . [the following:] On top of existing tariffs, the president imposed a 9% tariff on *all* imports in the first year (declining to 6% and then 3% in years two and three), and

a whopping 40% tariff on imports above last year's levels (dropping to 32% and 24%). . . .

The American Sheep Industry Association's president happily announced that the move will "bring some stability to the market." Whenever producers speak of bringing stability to the market, you know that consumers are getting fleeced.

The lamb decision, while little noticed at home, has been closely followed abroad. The decision undercuts the administration's free-trade rhetoric and harms its efforts to get other countries to open up their markets. Some import relief had been expected, but not so clearly protectionist as what finally materialized. The extreme decision has outraged farmers in Australia and New Zealand, and officials there have vowed to take the U.S. to a WTO dispute settlement panel.

The administration's timing could not have been worse. The decision came right after an Asia Pacific Economic Cooperation summit reaffirmed its commit-

ment to reduce trade barriers, and a few months before the World Trade Organization's November meeting in Seattle, where the WTO is to launch a new round of multilateral trade negotiations. A principal U.S. objective at the summit is the reduction of agricultural protection in Europe and elsewhere.

In 1947, facing an election the next year, President Truman courageously resisted special interest pressure and vetoed a bill to impose import quotas on wool, which would have jeopardized the first postwar multilateral trade negotiations due to start later that year. In contrast, Mr. Clinton, though a lame duck, caved in to political pressure. If the U.S., whose booming economy is the envy of the world, cannot resist protectionism, how can it expect other countries to do so?

SOURCE: *The Wall Street Journal*, July 12, 1999, p. A28.

To see how countries can benefit from trade, suppose there are two countries, the United States and Japan, and two goods, food and cars. Imagine that the two countries produce cars equally well: An American worker and a Japanese worker can each produce 1 car per month. By contrast, because the United States has more and better land, it is better at producing food: A U.S. worker can produce 2 tons of food per month, whereas a Japanese worker can produce only 1 ton of food per month.

The principle of comparative advantage states that each good should be produced by the country that has the smaller opportunity cost of producing that good. Because the opportunity cost of a car is 2 tons of food in the United States but only 1 ton of food in Japan, Japan has a comparative advantage in producing cars. Japan should produce more cars than it wants for its own use and export some of them to the United States. Similarly, because the opportunity cost of a ton of food is 1 car in Japan but only 1/2 car in the United States, the United States has a comparative advantage in producing food. The United States should produce more food than it wants to consume and export some of it to Japan. Through specialization and trade, both countries can have more food and more cars.

In reality, of course, the issues involved in trade among nations are more complex than this example suggests, as we will see in Chapter 9. Most important among these issues is that each country has many citizens with different interests. International trade can make some individuals worse off, even as it makes the country as a whole better off. When the United States exports food and imports cars, the impact on an American farmer is not the same as the impact on an American autoworker. Yet, contrary to the opinions sometimes voiced by politicians and political commentators, international trade is not like war, in which some countries win and others lose. Trade allows all countries to achieve greater prosperity.

QUICK QUIZ: Suppose that the world's fastest typist happens to be trained in brain surgery. Should he do his own typing or hire a secretary? Explain.

CONCLUSION

The principle of comparative advantage shows that trade can make everyone better off. You should now understand more fully the benefits of living in an interdependent economy. But having seen why interdependence is desirable, you might naturally ask how it is possible. How do free societies coordinate the diverse activities of all the people involved in their economies? What ensures that goods and services will get from those who should be producing them to those who should be consuming them?

In a world with only two people, such as the rancher and the farmer, the answer is simple: These two people can directly bargain and allocate resources between themselves. In the real world with billions of people, the answer is less obvious. We take up this issue in the next chapter, where we see that free societies allocate resources through the market forces of supply and demand.

Summary

- ◆ Each person consumes goods and services produced by many other people both in our country and around the world. Interdependence and trade are desirable because they allow everyone to enjoy a greater quantity and variety of goods and services.
- ◆ There are two ways to compare the ability of two people in producing a good. The person who can produce the good with the smaller quantity of inputs is said to have an *absolute advantage* in producing the good. The person who has the smaller opportunity cost of producing the good is said to have a *comparative advantage*. The gains from trade are based on comparative advantage, not absolute advantage.
- ◆ Trade makes everyone better off because it allows people to specialize in those activities in which they have a comparative advantage.
- ◆ The principle of comparative advantage applies to countries as well as to people. Economists use the principle of comparative advantage to advocate free trade among countries.

Key Concepts

absolute advantage, p. 53
opportunity cost, p. 53

comparative advantage, p. 53
imports, p. 57

exports, p. 57

Questions for Review

1. Explain how absolute advantage and comparative advantage differ.
2. Give an example in which one person has an absolute advantage in doing something but another person has a comparative advantage.
3. Is absolute advantage or comparative advantage more important for trade? Explain your reasoning, using the example in your answer to Question 2.
4. Will a nation tend to export or import goods for which it has a comparative advantage? Explain.
5. Why do economists oppose policies that restrict trade among nations?

Problems and Applications

1. Consider the farmer and the rancher from our example in this chapter. Explain why the farmer's opportunity cost of producing 1 pound of meat is 2 pounds of potatoes. Explain why the rancher's opportunity cost of producing 1 pound of meat is $1/8$ pound of potatoes.
2. Maria can read 20 pages of economics in an hour. She can also read 50 pages of sociology in an hour. She spends 5 hours per day studying.
 - a. Draw Maria's production possibilities frontier for reading economics and sociology.
 - b. What is Maria's opportunity cost of reading 100 pages of sociology?
3. American and Japanese workers can each produce 4 cars a year. An American worker can produce 10 tons of grain a year, whereas a Japanese worker can produce 5 tons of grain a year. To keep things simple, assume that each country has 100 million workers.
 - a. For this situation, construct a table analogous to Table 3-1.
 - b. Graph the production possibilities frontier of the American and Japanese economies.
 - c. For the United States, what is the opportunity cost of a car? Of grain? For Japan, what is the opportunity cost of a car? Of grain? Put

this information in a table analogous to Table 3-3.

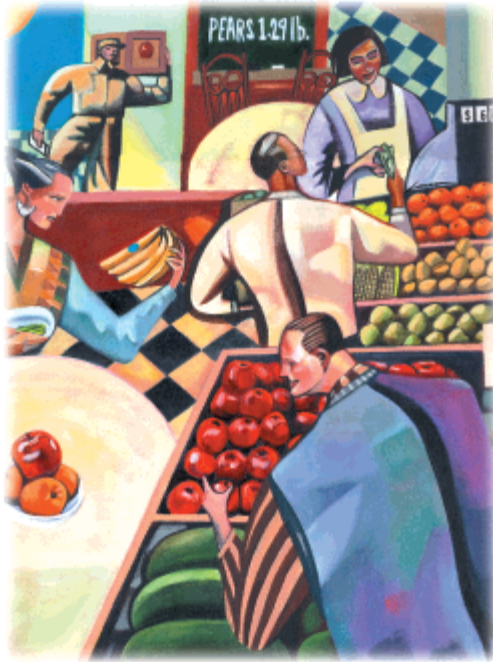
- d. Which country has an absolute advantage in producing cars? In producing grain?
 - e. Which country has a comparative advantage in producing cars? In producing grain?
 - f. Without trade, half of each country's workers produce cars and half produce grain. What quantities of cars and grain does each country produce?
 - g. Starting from a position without trade, give an example in which trade makes each country better off.
4. Pat and Kris are roommates. They spend most of their time studying (of course), but they leave some time for their favorite activities: making pizza and brewing root beer. Pat takes 4 hours to brew a gallon of root beer and 2 hours to make a pizza. Kris takes 6 hours to brew a gallon of root beer and 4 hours to make a pizza.
 - a. What is each roommate's opportunity cost of making a pizza? Who has the absolute advantage in making pizza? Who has the comparative advantage in making pizza?
 - b. If Pat and Kris trade foods with each other, who will trade away pizza in exchange for root beer?
 - c. The price of pizza can be expressed in terms of gallons of root beer. What is the highest price at which pizza can be traded that would make both roommates better off? What is the lowest price? Explain.
 5. Suppose that there are 10 million workers in Canada, and that each of these workers can produce either 2 cars or 30 bushels of wheat in a year.
 - a. What is the opportunity cost of producing a car in Canada? What is the opportunity cost of producing a bushel of wheat in Canada? Explain the relationship between the opportunity costs of the two goods.
 - b. Draw Canada's production possibilities frontier. If Canada chooses to consume 10 million cars, how much wheat can it consume without trade? Label this point on the production possibilities frontier.
 - c. Now suppose that the United States offers to buy 10 million cars from Canada in exchange for 20 bushels of wheat per car. If Canada continues to consume 10 million cars, how much wheat does this deal allow Canada to consume? Label this point on your diagram. Should Canada accept the deal?
 6. Consider a professor who is writing a book. The professor can both write the chapters and gather the needed data faster than anyone else at his university. Still, he pays a student to collect data at the library. Is this sensible? Explain.
 7. England and Scotland both produce scones and sweaters. Suppose that an English worker can produce 50 scones per hour or 1 sweater per hour. Suppose that a Scottish worker can produce 40 scones per hour or 2 sweaters per hour.
 - a. Which country has the absolute advantage in the production of each good? Which country has the comparative advantage?
 - b. If England and Scotland decide to trade, which commodity will Scotland trade to England? Explain.
 - c. If a Scottish worker could produce only 1 sweater per hour, would Scotland still gain from trade? Would England still gain from trade? Explain.
 8. Consider once again the farmer and rancher discussed in the chapter.
 - a. Suppose that a technological advance makes the farmer better at producing meat, so that he now needs only 2 hours to produce 1 pound of meat. What is his opportunity cost of meat and potatoes now? Does this alter his comparative advantage?
 - b. Is the deal that the rancher proposes—3 pounds of meat for 1 pound of potatoes—still good for the farmer? Explain.
 - c. Propose another deal to which the farmer and rancher might agree now.
 9. The following table describes the production possibilities of two cities in the country of Baseballia:

	PAIRS OF RED SOCKS PER WORKER PER HOUR	PAIRS OF WHITE SOCKS PER WORKER PER HOUR
BOSTON	3	3
CHICAGO	2	1

- a. Without trade, what is the price of white socks (in terms of red socks) in Boston? What is the price in Chicago?
- b. Which city has an absolute advantage in the production of each color sock? Which city has a comparative advantage in the production of each color sock?
- c. If the cities trade with each other, which color sock will each export?

- d. What is the range of prices at which trade can occur?
10. Suppose that all goods can be produced with fewer worker hours in Germany than in France.
- a. In what sense is the cost of all goods lower in Germany than in France?
 - b. In what sense is the cost of some goods lower in France?
 - c. If Germany and France traded with each other, would both countries be better off as a result? Explain in the context of your answers to parts (a) and (b).
11. Are the following statements true or false? Explain in each case.
- a. "Two countries can achieve gains from trade even if one of the countries has an absolute advantage in the production of all goods."
 - b. "Certain very talented people have a comparative advantage in everything they do."
 - c. "If a certain trade is good for one person, it can't be good for the other one."

4



THE MARKET FORCES OF SUPPLY AND DEMAND

When a cold snap hits Florida, the price of orange juice rises in supermarkets throughout the country. When the weather turns warm in New England every summer, the price of hotel rooms in the Caribbean plummets. When a war breaks out in the Middle East, the price of gasoline in the United States rises, and the price of a used Cadillac falls. What do these events have in common? They all show the workings of supply and demand.

Supply and demand are the two words that economists use most often—and for good reason. Supply and demand are the forces that make market economies work. They determine the quantity of each good produced and the price at which it is sold. If you want to know how any event or policy will affect the economy, you must think first about how it will affect supply and demand.

This chapter introduces the theory of supply and demand. It considers how buyers and sellers behave and how they interact with one another. It shows how

**IN THIS CHAPTER
YOU WILL . . .**

***Learn the nature of
a competitive
market***

***Examine what
determines the
demand for a good
in a competitive
market***

***Examine what
determines the
supply of a good in a
competitive market***

***See how supply and
demand together set
the price of a good
and the quantity
sold***

***Consider the key
role of prices in
allocating scarce
resources in market
economies***

supply and demand determine prices in a market economy and how prices, in turn, allocate the economy's scarce resources.

MARKETS AND COMPETITION

market

a group of buyers and sellers of a particular good or service

The terms *supply* and *demand* refer to the behavior of people as they interact with one another in markets. A **market** is a group of buyers and sellers of a particular good or service. The buyers as a group determine the demand for the product, and the sellers as a group determine the supply of the product. Before discussing how buyers and sellers behave, let's first consider more fully what we mean by a "market" and the various types of markets we observe in the economy.

COMPETITIVE MARKETS

Markets take many forms. Sometimes markets are highly organized, such as the markets for many agricultural commodities. In these markets, buyers and sellers meet at a specific time and place, where an auctioneer helps set prices and arrange sales.

More often, markets are less organized. For example, consider the market for ice cream in a particular town. Buyers of ice cream do not meet together at any one time. The sellers of ice cream are in different locations and offer somewhat different products. There is no auctioneer calling out the price of ice cream. Each seller posts a price for an ice-cream cone, and each buyer decides how much ice cream to buy at each store.

Even though it is not organized, the group of ice-cream buyers and ice-cream sellers forms a market. Each buyer knows that there are several sellers from which to choose, and each seller is aware that his product is similar to that offered by other sellers. The price of ice cream and the quantity of ice cream sold are not determined by any single buyer or seller. Rather, price and quantity are determined by all buyers and sellers as they interact in the marketplace.

The market for ice cream, like most markets in the economy, is highly competitive. A **competitive market** is a market in which there are many buyers and many sellers so that each has a negligible impact on the market price. Each seller of ice cream has limited control over the price because other sellers are offering similar products. A seller has little reason to charge less than the going price, and if he or she charges more, buyers will make their purchases elsewhere. Similarly, no single buyer of ice cream can influence the price of ice cream because each buyer purchases only a small amount.

In this chapter we examine how buyers and sellers interact in competitive markets. We see how the forces of supply and demand determine both the quantity of the good sold and its price.

COMPETITION: PERFECT AND OTHERWISE

We assume in this chapter that markets are *perfectly competitive*. Perfectly competitive markets are defined by two primary characteristics: (1) the goods being offered for sale are all the same, and (2) the buyers and sellers are so numerous that

competitive market

a market in which there are many buyers and many sellers so that each has a negligible impact on the market price

no single buyer or seller can influence the market price. Because buyers and sellers in perfectly competitive markets must accept the price the market determines, they are said to be *price takers*.

There are some markets in which the assumption of perfect competition applies perfectly. In the wheat market, for example, there are thousands of farmers who sell wheat and millions of consumers who use wheat and wheat products. Because no single buyer or seller can influence the price of wheat, each takes the price as given.

Not all goods and services, however, are sold in perfectly competitive markets. Some markets have only one seller, and this seller sets the price. Such a seller is called a *monopoly*. Your local cable television company, for instance, may be a monopoly. Residents of your town probably have only one cable company from which to buy this service.

Some markets fall between the extremes of perfect competition and monopoly. One such market, called an *oligopoly*, has a few sellers that do not always compete aggressively. Airline routes are an example. If a route between two cities is serviced by only two or three carriers, the carriers may avoid rigorous competition to keep prices high. Another type of market is *monopolistically competitive*; it contains many sellers, each offering a slightly different product. Because the products are not exactly the same, each seller has some ability to set the price for its own product. An example is the software industry. Many word processing programs compete with one another for users, but every program is different from every other and has its own price.

Despite the diversity of market types we find in the world, we begin by studying perfect competition. Perfectly competitive markets are the easiest to analyze. Moreover, because some degree of competition is present in most markets, many of the lessons that we learn by studying supply and demand under perfect competition apply in more complicated markets as well.

QUICK QUIZ: What is a market? ♦ What does it mean for a market to be competitive?

DEMAND

We begin our study of markets by examining the behavior of buyers. Here we consider what determines the **quantity demanded** of any good, which is the amount of the good that buyers are willing and able to purchase. To focus our thinking, let's keep in mind a particular good—ice cream.

quantity demanded

the amount of a good that buyers are willing and able to purchase

WHAT DETERMINES THE QUANTITY AN INDIVIDUAL DEMANDS?

Consider your own demand for ice cream. How do you decide how much ice cream to buy each month, and what factors affect your decision? Here are some of the answers you might give.

law of demand

the claim that, other things equal, the quantity demanded of a good falls when the price of the good rises

normal good

a good for which, other things equal, an increase in income leads to an increase in demand

inferior good

a good for which, other things equal, an increase in income leads to a decrease in demand

substitutes

two goods for which an increase in the price of one leads to an increase in the demand for the other

complements

two goods for which an increase in the price of one leads to a decrease in the demand for the other

Price If the price of ice cream rose to \$20 per scoop, you would buy less ice cream. You might buy frozen yogurt instead. If the price of ice cream fell to \$0.20 per scoop, you would buy more. Because the quantity demanded falls as the price rises and rises as the price falls, we say that the quantity demanded is *negatively related* to the price. This relationship between price and quantity demanded is true for most goods in the economy and, in fact, is so pervasive that economists call it the **law of demand**: Other things equal, when the price of a good rises, the quantity demanded of the good falls.

Income What would happen to your demand for ice cream if you lost your job one summer? Most likely, it would fall. A lower income means that you have less to spend in total, so you would have to spend less on some—and probably most—goods. If the demand for a good falls when income falls, the good is called a **normal good**.

Not all goods are normal goods. If the demand for a good rises when income falls, the good is called an **inferior good**. An example of an inferior good might be bus rides. As your income falls, you are less likely to buy a car or take a cab, and more likely to ride the bus.

Prices of Related Goods Suppose that the price of frozen yogurt falls. The law of demand says that you will buy more frozen yogurt. At the same time, you will probably buy less ice cream. Because ice cream and frozen yogurt are both cold, sweet, creamy desserts, they satisfy similar desires. When a fall in the price of one good reduces the demand for another good, the two goods are called **substitutes**. Substitutes are often pairs of goods that are used in place of each other, such as hot dogs and hamburgers, sweaters and sweatshirts, and movie tickets and video rentals.

Now suppose that the price of hot fudge falls. According to the law of demand, you will buy more hot fudge. Yet, in this case, you will buy *more* ice cream as well, because ice cream and hot fudge are often used together. When a fall in the price of one good raises the demand for another good, the two goods are called **complements**. Complements are often pairs of goods that are used together, such as gasoline and automobiles, computers and software, and skis and ski lift tickets.

Tastes The most obvious determinant of your demand is your tastes. If you like ice cream, you buy more of it. Economists normally do not try to explain people's tastes because tastes are based on historical and psychological forces that are beyond the realm of economics. Economists do, however, examine what happens when tastes change.

Expectations Your expectations about the future may affect your demand for a good or service today. For example, if you expect to earn a higher income next month, you may be more willing to spend some of your current savings buying ice cream. As another example, if you expect the price of ice cream to fall tomorrow, you may be less willing to buy an ice-cream cone at today's price.

THE DEMAND SCHEDULE AND THE DEMAND CURVE

We have seen that many variables determine the quantity of ice cream a person demands. Imagine that we hold all these variables constant except one—the price. Let’s consider how the price affects the quantity of ice cream demanded.

Table 4-1 shows how many ice-cream cones Catherine buys each month at different prices of ice cream. If ice cream is free, Catherine eats 12 cones. At \$0.50 per cone, Catherine buys 10 cones. As the price rises further, she buys fewer and fewer cones. When the price reaches \$3.00, Catherine doesn’t buy any ice cream at all. Table 4-1 is a **demand schedule**, a table that shows the relationship between the price of a good and the quantity demanded. (Economists use the term *schedule* because the table, with its parallel columns of numbers, resembles a train schedule.)

Figure 4-1 graphs the numbers in Table 4-1. By convention, the price of ice cream is on the vertical axis, and the quantity of ice cream demanded is on the

demand schedule
a table that shows the relationship between the price of a good and the quantity demanded

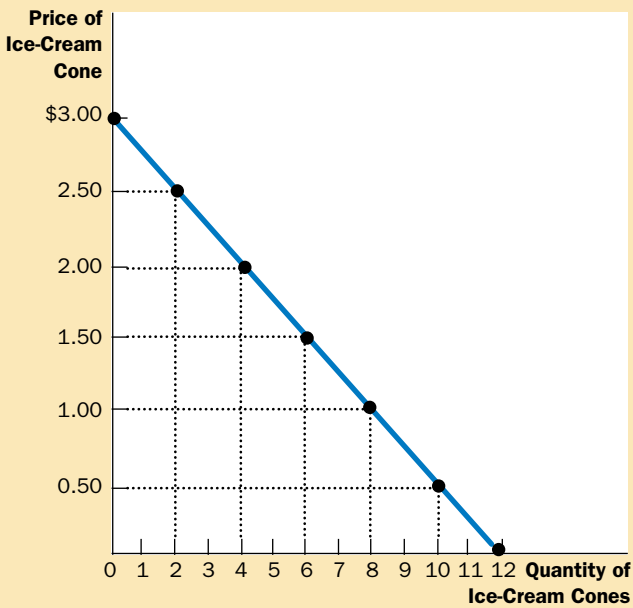
Table 4-1

CATHERINE’S DEMAND SCHEDULE. The demand schedule shows the quantity demanded at each price.

PRICE OF ICE-CREAM CONE	QUANTITY OF CONES DEMANDED
\$0.00	12
0.50	10
1.00	8
1.50	6
2.00	4
2.50	2
3.00	0

Figure 4-1

CATHERINE’S DEMAND CURVE. This demand curve, which graphs the demand schedule in Table 4-1, shows how the quantity demanded of the good changes as its price varies. Because a lower price increases the quantity demanded, the demand curve slopes downward.



demand curve

a graph of the relationship between the price of a good and the quantity demanded

ceteris paribus

a Latin phrase, translated as “other things being equal,” used as a reminder that all variables other than the ones being studied are assumed to be constant

horizontal axis. The downward-sloping line relating price and quantity demanded is called the **demand curve**.

CETERIS PARIBUS

Whenever you see a demand curve, remember that it is drawn holding many things constant. Catherine’s demand curve in Figure 4-1 shows what happens to the quantity of ice cream Catherine demands when only the price of ice cream varies. The curve is drawn assuming that Catherine’s income, tastes, expectations, and the prices of related products are not changing.

Economists use the term *ceteris paribus* to signify that all the relevant variables, except those being studied at that moment, are held constant. The Latin phrase literally means “other things being equal.” The demand curve slopes downward because, *ceteris paribus*, lower prices mean a greater quantity demanded.

Although the term *ceteris paribus* refers to a hypothetical situation in which some variables are assumed to be constant, in the real world many things change at the same time. For this reason, when we use the tools of supply and demand to analyze events or policies, it is important to keep in mind what is being held constant and what is not.

MARKET DEMAND VERSUS INDIVIDUAL DEMAND

So far we have talked about an individual’s demand for a product. To analyze how markets work, we need to determine the *market demand*, which is the sum of all the individual demands for a particular good or service.

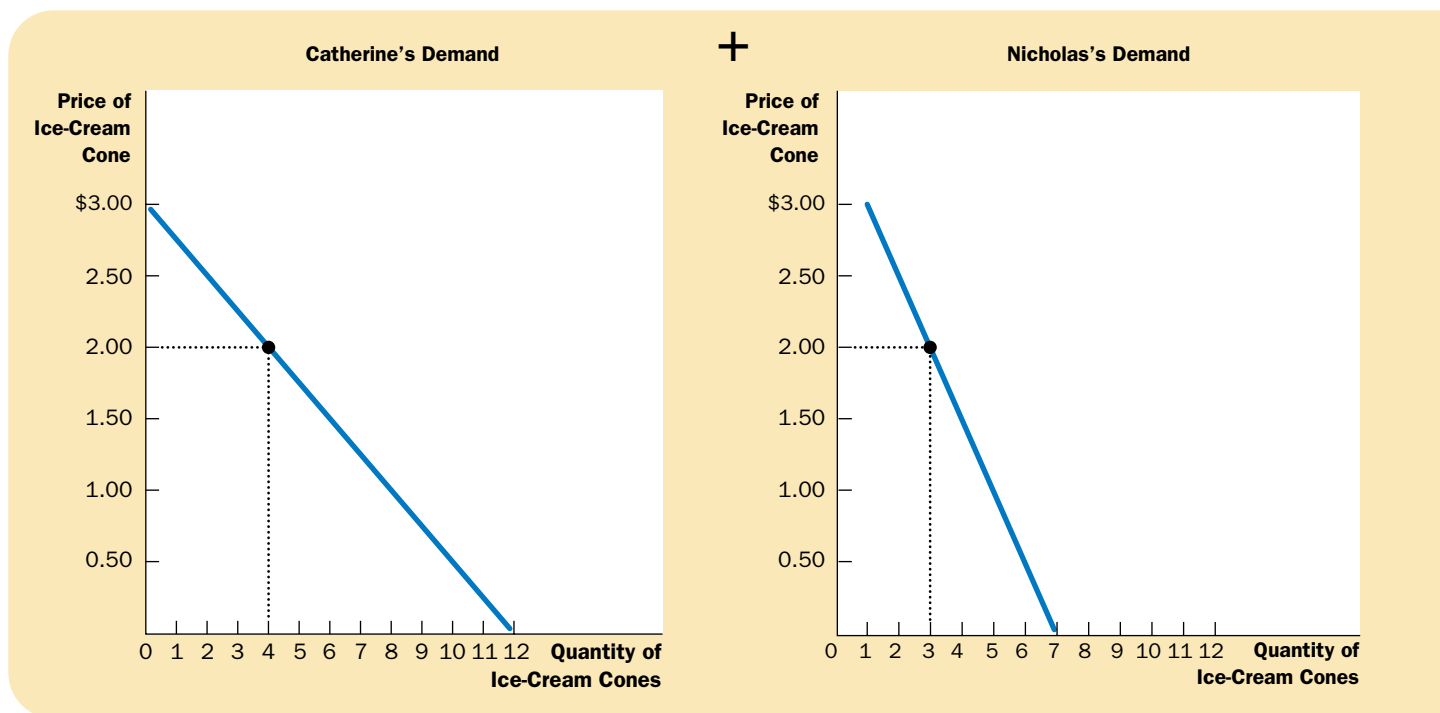


Table 4-2 shows the demand schedules for ice cream of two individuals—Catherine and Nicholas. At any price, Catherine’s demand schedule tells us how much ice cream she buys, and Nicholas’s demand schedule tells us how much ice cream he buys. The market demand is the sum of the two individual demands.

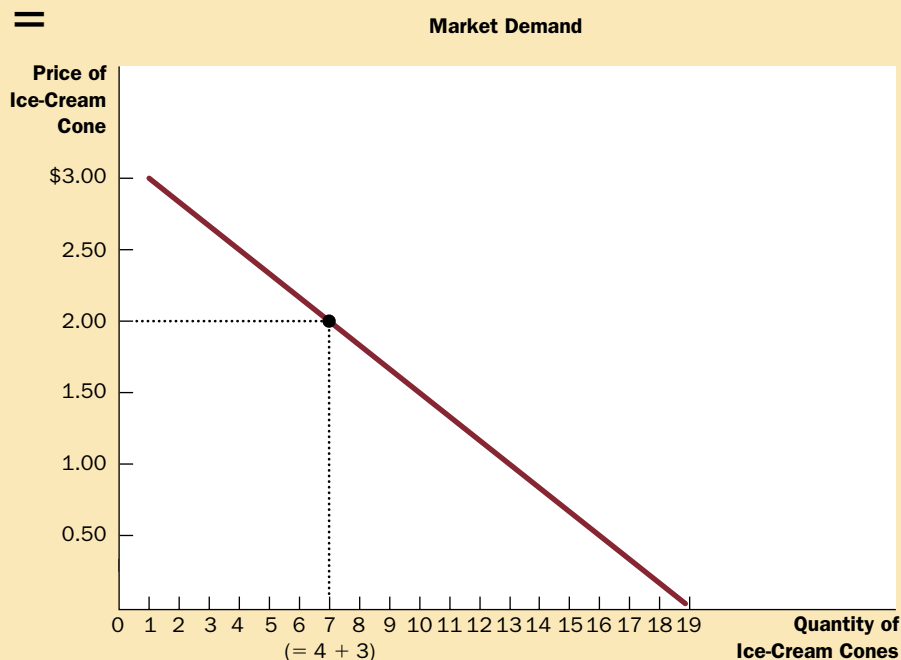
Because market demand is derived from individual demands, it depends on all those factors that determine the demand of individual buyers. Thus, market demand depends on buyers’ incomes, tastes, expectations, and the prices of related goods. It also depends on the number of buyers. (If Peter, another consumer of ice cream, were to join Catherine and Nicholas, the quantity demanded in the market would be higher at every price.) The demand schedules in Table 4-2 show what happens to quantity demanded as the price varies while all the other variables that determine quantity demanded are held constant.

Figure 4-2 shows the demand curves that correspond to these demand schedules. Notice that we sum the individual demand curves *horizontally* to obtain the

Table 4-2

PRICE OF ICE-CREAM CONE	CATHERINE		NICHOLAS		MARKET
\$0.00	12	+	7	=	19
0.50	10		6		16
1.00	8		5		13
1.50	6		4		10
2.00	4		3		7
2.50	2		2		4
3.00	0		1		1

INDIVIDUAL AND MARKET DEMAND SCHEDULES. The quantity demanded in a market is the sum of the quantities demanded by all the buyers.

**Figure 4-2**

MARKET DEMAND AS THE SUM OF INDIVIDUAL DEMANDS. The market demand curve is found by adding horizontally the individual demand curves. At a price of \$2, Catherine demands 4 ice-cream cones, and Nicholas demands 3 ice-cream cones. The quantity demanded in the market at this price is 7 cones.

market demand curve. That is, to find the total quantity demanded at any price, we add the individual quantities found on the horizontal axis of the individual demand curves. Because we are interested in analyzing how markets work, we will work most often with the market demand curve. The market demand curve shows how the total quantity demanded of a good varies as the price of the good varies.

SHIFTS IN THE DEMAND CURVE

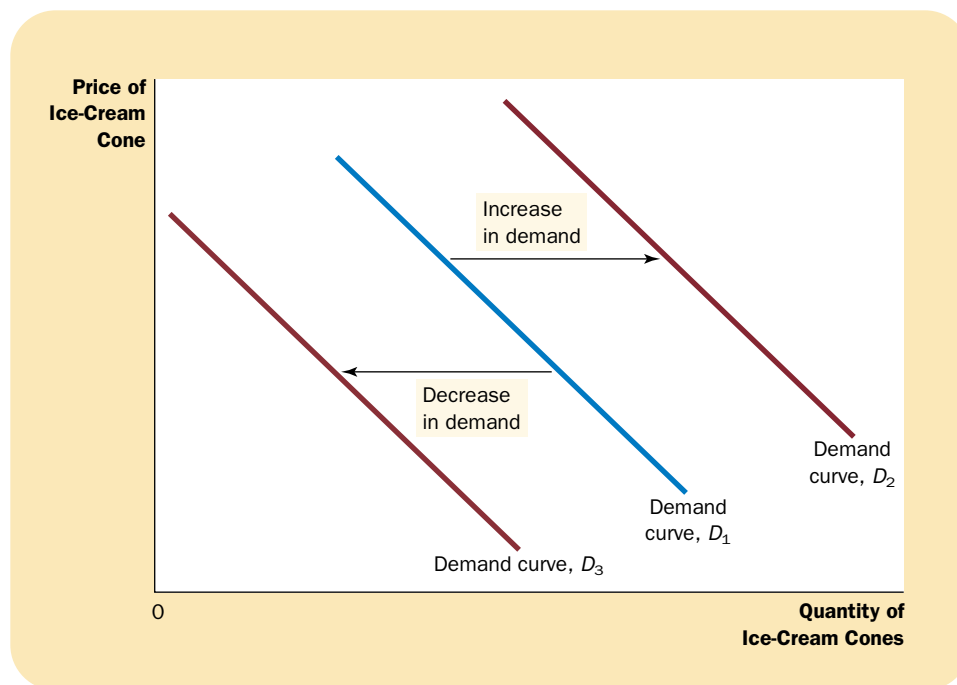
Suppose that the American Medical Association suddenly announces a new discovery: People who regularly eat ice cream live longer, healthier lives. How does this announcement affect the market for ice cream? The discovery changes people's tastes and raises the demand for ice cream. At any given price, buyers now want to purchase a larger quantity of ice cream, and the demand curve for ice cream shifts to the right.

Whenever any determinant of demand changes, other than the good's price, the demand curve shifts. As Figure 4-3 shows, any change that increases the quantity demanded at every price shifts the demand curve to the right. Similarly, any change that reduces the quantity demanded at every price shifts the demand curve to the left.

Table 4-3 lists the variables that determine the quantity demanded in a market and how a change in the variable affects the demand curve. Notice that price plays a special role in this table. Because price is on the vertical axis when we graph a demand curve, a change in price does not shift the curve but represents a movement along it. By contrast, when there is a change in income, the prices of related goods, tastes, expectations, or the number of buyers, the quantity demanded at each price changes; this is represented by a shift in the demand curve.

Figure 4-3

SHIFTS IN THE DEMAND CURVE. Any change that raises the quantity that buyers wish to purchase at a given price shifts the demand curve to the right. Any change that lowers the quantity that buyers wish to purchase at a given price shifts the demand curve to the left.



**VARIABLES THAT AFFECT
QUANTITY DEMANDED**
A CHANGE IN THIS VARIABLE . . .

Price	Represents a movement along the demand curve
Income	Shifts the demand curve
Prices of related goods	Shifts the demand curve
Tastes	Shifts the demand curve
Expectations	Shifts the demand curve
Number of buyers	Shifts the demand curve

Table 4-3

THE DETERMINANTS OF QUANTITY DEMANDED. This table lists the variables that can influence the quantity demanded in a market. Notice the special role that price plays: A change in the price represents a movement along the demand curve, whereas a change in one of the other variables shifts the demand curve.

In summary, *the demand curve shows what happens to the quantity demanded of a good when its price varies, holding constant all other determinants of quantity demanded. When one of these other determinants changes, the demand curve shifts.*

CASE STUDY TWO WAYS TO REDUCE THE QUANTITY OF SMOKING DEMANDED

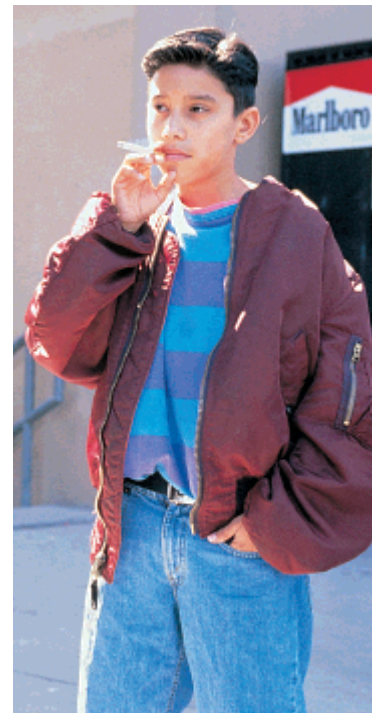
Public policymakers often want to reduce the amount that people smoke. There are two ways that policy can attempt to achieve this goal.

One way to reduce smoking is to shift the demand curve for cigarettes and other tobacco products. Public service announcements, mandatory health warnings on cigarette packages, and the prohibition of cigarette advertising on television are all policies aimed at reducing the quantity of cigarettes demanded at any given price. If successful, these policies shift the demand curve for cigarettes to the left, as in panel (a) of Figure 4-4.

Alternatively, policymakers can try to raise the price of cigarettes. If the government taxes the manufacture of cigarettes, for example, cigarette companies pass much of this tax on to consumers in the form of higher prices. A higher price encourages smokers to reduce the numbers of cigarettes they smoke. In this case, the reduced amount of smoking does not represent a shift in the demand curve. Instead, it represents a movement along the same demand curve to a point with a higher price and lower quantity, as in panel (b) of Figure 4-4.

How much does the amount of smoking respond to changes in the price of cigarettes? Economists have attempted to answer this question by studying what happens when the tax on cigarettes changes. They have found that a 10 percent increase in the price causes a 4 percent reduction in the quantity demanded. Teenagers are found to be especially sensitive to the price of cigarettes: A 10 percent increase in the price causes a 12 percent drop in teenage smoking.

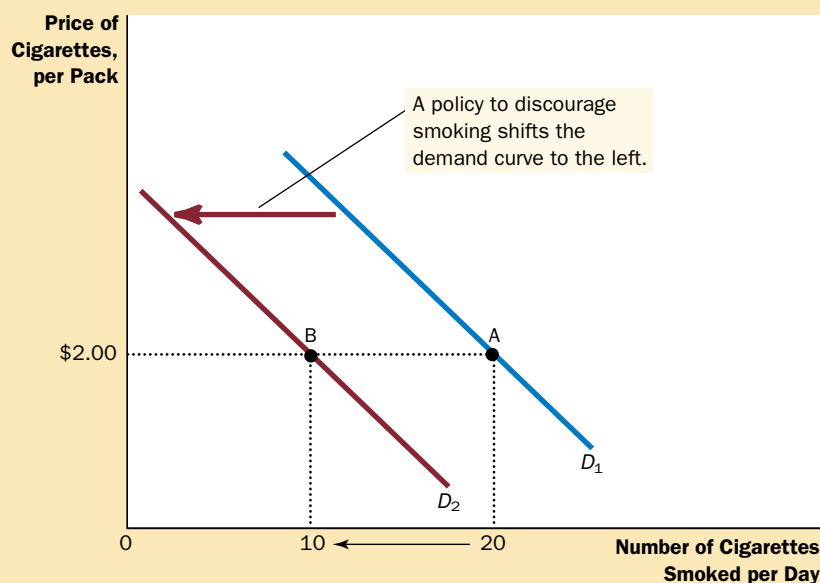
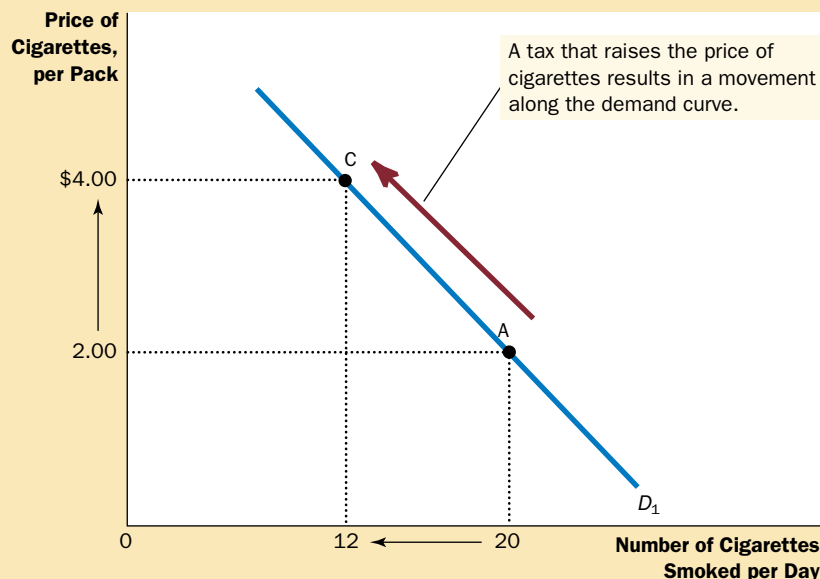
A related question is how the price of cigarettes affects the demand for illicit drugs, such as marijuana. Opponents of cigarette taxes often argue that tobacco and marijuana are substitutes, so that high cigarette prices encourage marijuana use. By contrast, many experts on substance abuse view tobacco as a “gateway drug” leading the young to experiment with other harmful substances. Most studies of the data are consistent with this view: They find that lower cigarette prices are associated with greater use of marijuana. In other words, tobacco and marijuana appear to be complements rather than substitutes.



WHAT IS THE BEST WAY TO STOP THIS?

Figure 4-4

SHIFTS IN THE DEMAND CURVE VERSUS MOVEMENTS ALONG THE DEMAND CURVE. If warnings on cigarette packages convince smokers to smoke less, the demand curve for cigarettes shifts to the left. In panel (a), the demand curve shifts from D_1 to D_2 . At a price of \$2 per pack, the quantity demanded falls from 20 to 10 cigarettes per day, as reflected by the shift from point A to point B. By contrast, if a tax raises the price of cigarettes, the demand curve does not shift. Instead, we observe a movement to a different point on the demand curve. In panel (b), when the price rises from \$2 to \$4, the quantity demanded falls from 20 to 12 cigarettes per day, as reflected by the movement from point A to point C.

(a) A Shift in the Demand Curve**(b) A Movement along the Demand Curve**

QUICK QUIZ: List the determinants of the quantity of pizza you demand.

- ◆ Make up an example of a demand schedule for pizza, and graph the implied demand curve.
- ◆ Give an example of something that would shift this demand curve.
- ◆ Would a change in the price of pizza shift this demand curve?

SUPPLY

We now turn to the other side of the market and examine the behavior of sellers. The **quantity supplied** of any good or service is the amount that sellers are willing and able to sell. Once again, to focus our thinking, let's consider the market for ice cream and look at the factors that determine the quantity supplied.

quantity supplied

the amount of a good that sellers are willing and able to sell

WHAT DETERMINES THE QUANTITY AN INDIVIDUAL SUPPLIES?

Imagine that you are running Student Sweets, a company that produces and sells ice cream. What determines the quantity of ice cream you are willing to produce and offer for sale? Here are some possible answers.

Price The price of ice cream is one determinant of the quantity supplied. When the price of ice cream is high, selling ice cream is profitable, and so the quantity supplied is large. As a seller of ice cream, you work long hours, buy many ice-cream machines, and hire many workers. By contrast, when the price of ice cream is low, your business is less profitable, and so you will produce less ice cream. At an even lower price, you may choose to go out of business altogether, and your quantity supplied falls to zero.

Because the quantity supplied rises as the price rises and falls as the price falls, we say that the quantity supplied is *positively related* to the price of the good. This relationship between price and quantity supplied is called the **law of supply**: Other things equal, when the price of a good rises, the quantity supplied of the good also rises.

law of supply

the claim that, other things equal, the quantity supplied of a good rises when the price of the good rises

Input Prices To produce its output of ice cream, Student Sweets uses various inputs: cream, sugar, flavoring, ice-cream machines, the buildings in which the ice cream is made, and the labor of workers to mix the ingredients and operate the machines. When the price of one or more of these inputs rises, producing ice cream is less profitable, and your firm supplies less ice cream. If input prices rise substantially, you might shut down your firm and supply no ice cream at all. Thus, the supply of a good is negatively related to the price of the inputs used to make the good.

Technology The technology for turning the inputs into ice cream is yet another determinant of supply. The invention of the mechanized ice-cream machine, for example, reduced the amount of labor necessary to make ice cream. By reducing firms' costs, the advance in technology raised the supply of ice cream.

Expectations The amount of ice cream you supply today may depend on your expectations of the future. For example, if you expect the price of ice cream to rise in the future, you will put some of your current production into storage and supply less to the market today.

THE SUPPLY SCHEDULE AND THE SUPPLY CURVE

supply schedule
a table that shows the relationship between the price of a good and the quantity supplied

supply curve
a graph of the relationship between the price of a good and the quantity supplied

Consider how the quantity supplied varies with the price, holding input prices, technology, and expectations constant. Table 4-4 shows the quantity supplied by Ben, an ice-cream seller, at various prices of ice cream. At a price below \$1.00, Ben does not supply any ice cream at all. As the price rises, he supplies a greater and greater quantity. This table is called the **supply schedule**.

Figure 4-5 graphs the relationship between the quantity of ice cream supplied and the price. The curve relating price and quantity supplied is called the **supply curve**. The supply curve slopes upward because, *ceteris paribus*, a higher price means a greater quantity supplied.

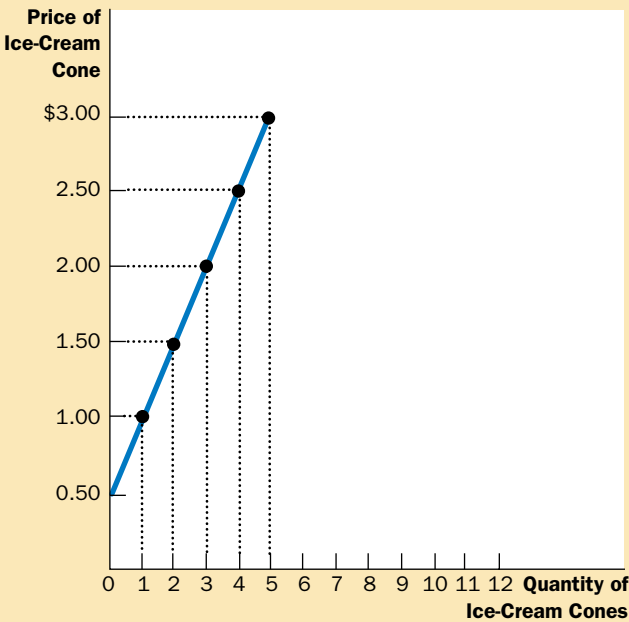
Table 4-4

BEN'S SUPPLY SCHEDULE. The supply schedule shows the quantity supplied at each price.

PRICE OF ICE-CREAM CONE	QUANTITY OF CONES SUPPLIED
\$0.00	0
0.50	0
1.00	1
1.50	2
2.00	3
2.50	4
3.00	5

Figure 4-5

BEN'S SUPPLY CURVE. This supply curve, which graphs the supply schedule in Table 4-4, shows how the quantity supplied of the good changes as its price varies. Because a higher price increases the quantity supplied, the supply curve slopes upward.



MARKET SUPPLY VERSUS INDIVIDUAL SUPPLY

Just as market demand is the sum of the demands of all buyers, market supply is the sum of the supplies of all sellers. Table 4-5 shows the supply schedules for two ice-cream producers—Ben and Jerry. At any price, Ben’s supply schedule tells us the quantity of ice cream Ben supplies, and Jerry’s supply schedule tells us the quantity of ice cream Jerry supplies. The market supply is the sum of the two individual supplies.

Market supply depends on all those factors that influence the supply of individual sellers, such as the prices of inputs used to produce the good, the available technology, and expectations. In addition, the supply in a market depends on the number of sellers. (If Ben or Jerry were to retire from the ice-cream business, the supply in the market would fall.) The supply schedules in Table 4-5 show what happens to quantity supplied as the price varies while all the other variables that determine quantity supplied are held constant.

Figure 4-6 shows the supply curves that correspond to the supply schedules in Table 4-5. As with demand curves, we sum the individual supply curves *horizontally* to obtain the market supply curve. That is, to find the total quantity supplied at any price, we add the individual quantities found on the horizontal axis of the individual supply curves. The market supply curve shows how the total quantity supplied varies as the price of the good varies.

SHIFTS IN THE SUPPLY CURVE

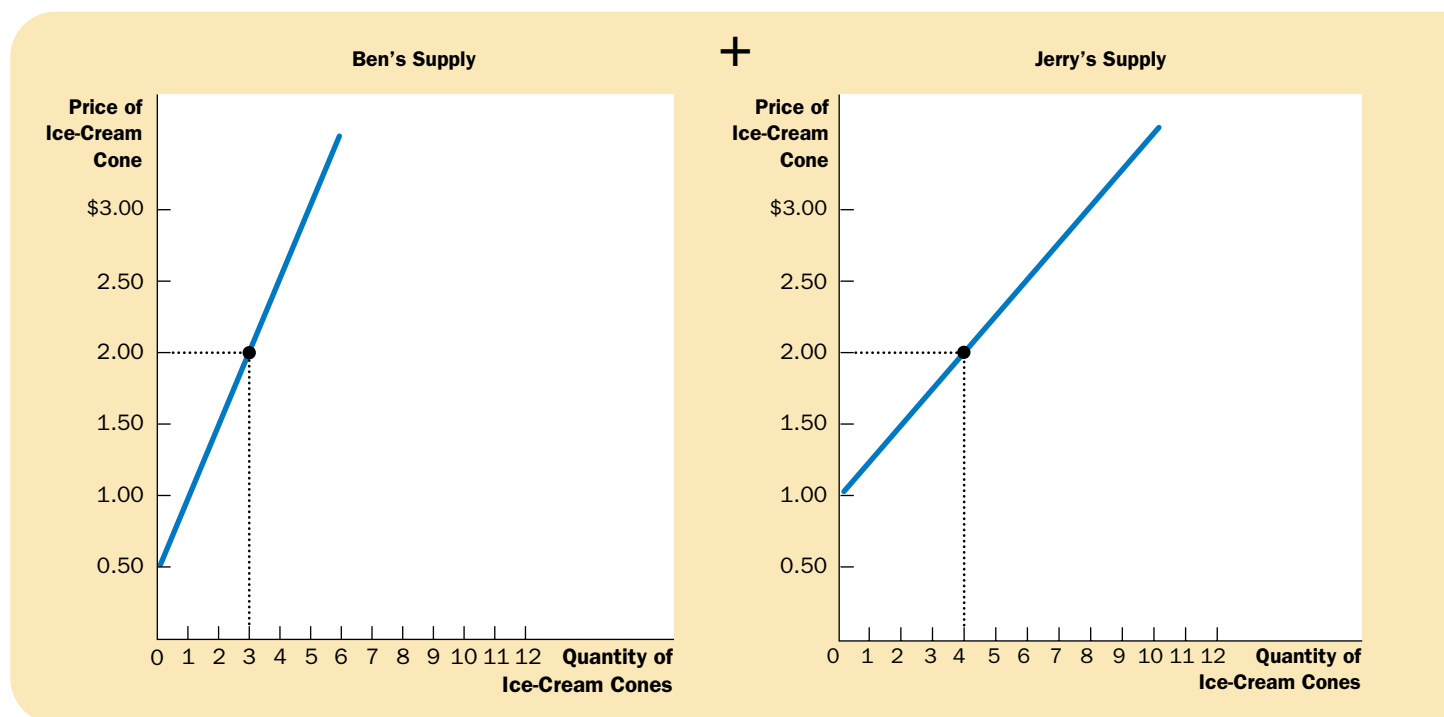
Suppose that the price of sugar falls. How does this change affect the supply of ice cream? Because sugar is an input into producing ice cream, the fall in the price of sugar makes selling ice cream more profitable. This raises the supply of ice cream: At any given price, sellers are now willing to produce a larger quantity. Thus, the supply curve for ice cream shifts to the right.

Whenever there is a change in any determinant of supply, other than the good’s price, the supply curve shifts. As Figure 4-7 shows, any change that raises quantity supplied at every price shifts the supply curve to the right. Similarly, any change that reduces the quantity supplied at every price shifts the supply curve to the left.

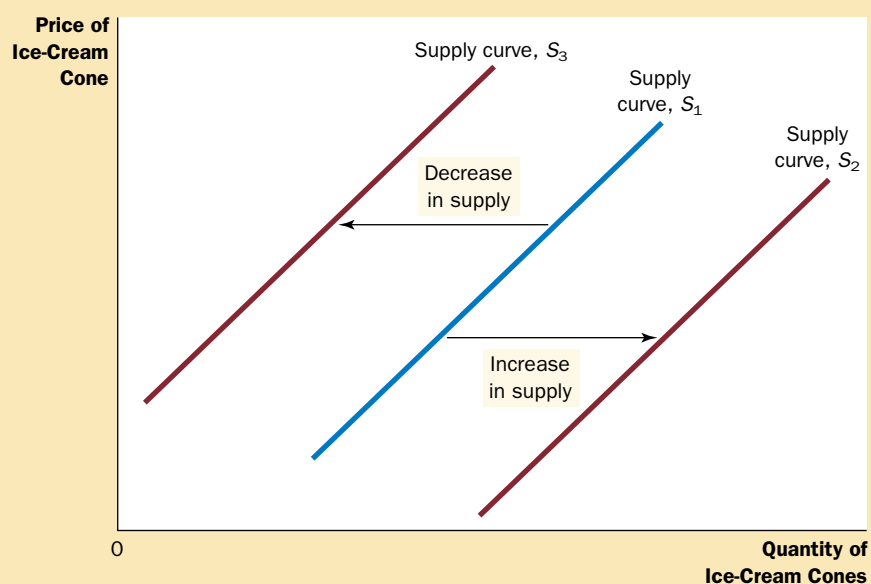
PRICE OF ICE-CREAM CONE	BEN		JERRY		MARKET
\$0.00	0	+	0	=	0
0.50	0		0		0
1.00	1		0		1
1.50	2		2		4
2.00	3		4		7
2.50	4		6		10
3.00	5		8		13

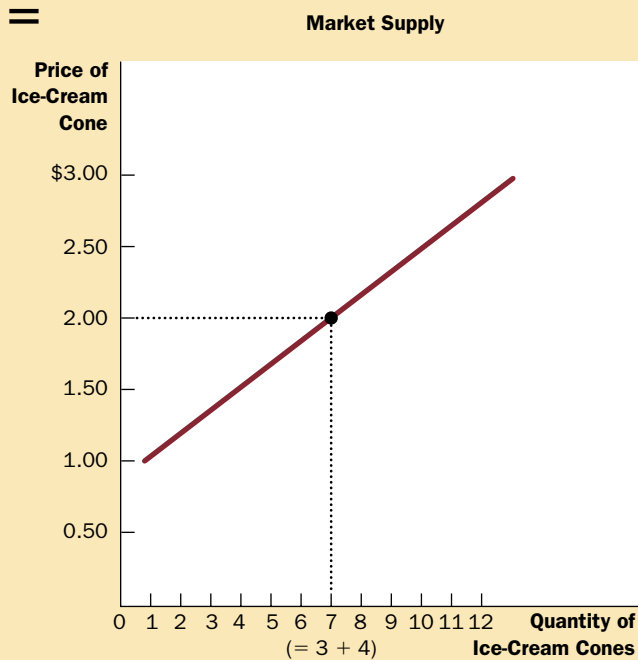
Table 4-5

INDIVIDUAL AND MARKET SUPPLY SCHEDULES. The quantity supplied in a market is the sum of the quantities supplied by all the sellers.

**Figure 4-7**

SHIFTS IN THE SUPPLY CURVE. Any change that raises the quantity that sellers wish to produce at a given price shifts the supply curve to the right. Any change that lowers the quantity that sellers wish to produce at a given price shifts the supply curve to the left.



**Figure 4-6**

MARKET SUPPLY AS THE SUM OF INDIVIDUAL SUPPLIES. The market supply curve is found by adding horizontally the individual supply curves. At a price of \$2, Ben supplies 3 ice-cream cones, and Jerry supplies 4 ice-cream cones. The quantity supplied in the market at this price is 7 cones.

**VARIABLES THAT AFFECT
QUANTITY SUPPLIED**

A CHANGE IN THIS VARIABLE . . .

Price	Represents a movement along the supply curve
Input prices	Shifts the supply curve
Technology	Shifts the supply curve
Expectations	Shifts the supply curve
Number of sellers	Shifts the supply curve

Table 4-6

THE DETERMINANTS OF QUANTITY SUPPLIED. This table lists the variables that can influence the quantity supplied in a market. Notice the special role that price plays: A change in the price represents a movement along the supply curve, whereas a change in one of the other variables shifts the supply curve.

Table 4-6 lists the variables that determine the quantity supplied in a market and how a change in the variable affects the supply curve. Once again, price plays a special role in the table. Because price is on the vertical axis when we graph a supply curve, a change in price does not shift the curve but represents a movement along it. By contrast, when there is a change in input prices, technology, expectations, or the number of sellers, the quantity supplied at each price changes; this is represented by a shift in the supply curve.

In summary, *the supply curve shows what happens to the quantity supplied of a good when its price varies, holding constant all other determinants of quantity supplied. When one of these other determinants changes, the supply curve shifts.*

QUICK QUIZ: List the determinants of the quantity of pizza supplied.

- ◆ Make up an example of a supply schedule for pizza, and graph the implied supply curve.
- ◆ Give an example of something that would shift this supply curve.
- ◆ Would a change in the price of pizza shift this supply curve?

SUPPLY AND DEMAND TOGETHER

Having analyzed supply and demand separately, we now combine them to see how they determine the quantity of a good sold in a market and its price.

equilibrium

a situation in which supply and demand have been brought into balance

equilibrium price

the price that balances supply and demand

equilibrium quantity

the quantity supplied and the quantity demanded when the price has adjusted to balance supply and demand

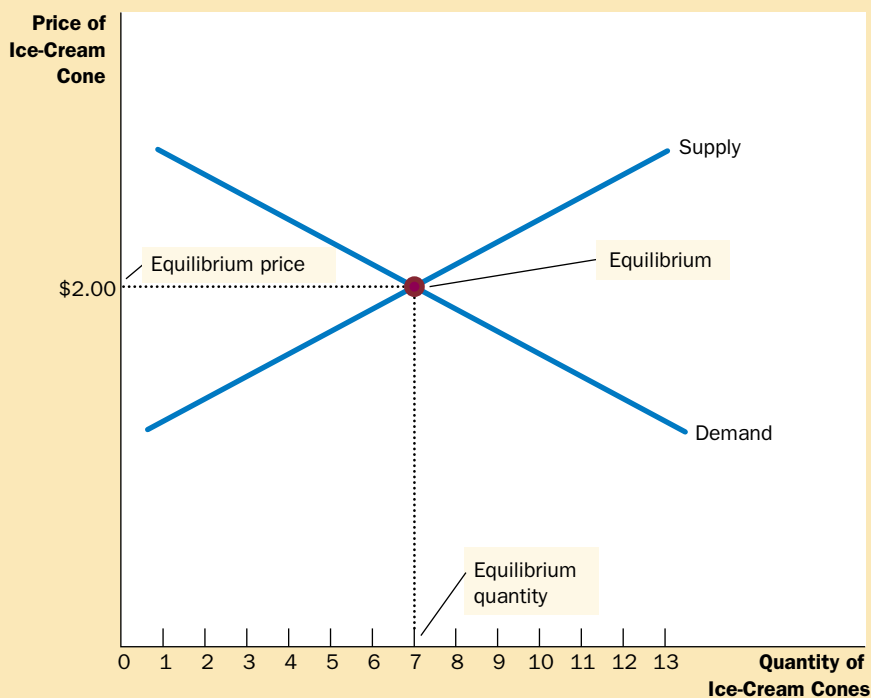
EQUILIBRIUM

Figure 4-8 shows the market supply curve and market demand curve together. Notice that there is one point at which the supply and demand curves intersect; this point is called the market's **equilibrium**. The price at which these two curves cross is called the **equilibrium price**, and the quantity is called the **equilibrium quantity**. Here the equilibrium price is \$2.00 per cone, and the equilibrium quantity is 7 ice-cream cones.

The dictionary defines the word *equilibrium* as a situation in which various forces are in balance—and this also describes a market's equilibrium. *At the*

Figure 4-8

THE EQUILIBRIUM OF SUPPLY AND DEMAND. The equilibrium is found where the supply and demand curves intersect. At the equilibrium price, the quantity supplied equals the quantity demanded. Here the equilibrium price is \$2: At this price, 7 ice-cream cones are supplied, and 7 ice-cream cones are demanded.



equilibrium price, the quantity of the good that buyers are willing and able to buy exactly balances the quantity that sellers are willing and able to sell. The equilibrium price is sometimes called the *market-clearing price* because, at this price, everyone in the market has been satisfied: Buyers have bought all they want to buy, and sellers have sold all they want to sell.

The actions of buyers and sellers naturally move markets toward the equilibrium of supply and demand. To see why, consider what happens when the market price is not equal to the equilibrium price.

Suppose first that the market price is above the equilibrium price, as in panel (a) of Figure 4-9. At a price of \$2.50 per cone, the quantity of the good supplied (10 cones) exceeds the quantity demanded (4 cones). There is a **surplus** of the good: Suppliers are unable to sell all they want at the going price. When there is a surplus in the ice-cream market, for instance, sellers of ice cream find their freezers increasingly full of ice cream they would like to sell but cannot. They respond to the surplus by cutting their prices. Prices continue to fall until the market reaches the equilibrium.

Suppose now that the market price is below the equilibrium price, as in panel (b) of Figure 4-9. In this case, the price is \$1.50 per cone, and the quantity of the good demanded exceeds the quantity supplied. There is a **shortage** of the good: Demanders are unable to buy all they want at the going price. When a shortage occurs in the ice-cream market, for instance, buyers have to wait in long lines for a chance to buy one of the few cones that are available. With too many buyers chasing too few goods, sellers can respond to the shortage by raising their prices without losing sales. As prices rise, the market once again moves toward the equilibrium.

Thus, the activities of the many buyers and sellers automatically push the market price toward the equilibrium price. Once the market reaches its equilibrium, all buyers and sellers are satisfied, and there is no upward or downward pressure on the price. How quickly equilibrium is reached varies from market to market, depending on how quickly prices adjust. In most free markets, however, surpluses and shortages are only temporary because prices eventually move toward their equilibrium levels. Indeed, this phenomenon is so pervasive that it is sometimes called the **law of supply and demand**: The price of any good adjusts to bring the supply and demand for that good into balance.

surplus

a situation in which quantity supplied is greater than quantity demanded

shortage

a situation in which quantity demanded is greater than quantity supplied

law of supply and demand

the claim that the price of any good adjusts to bring the supply and demand for that good into balance

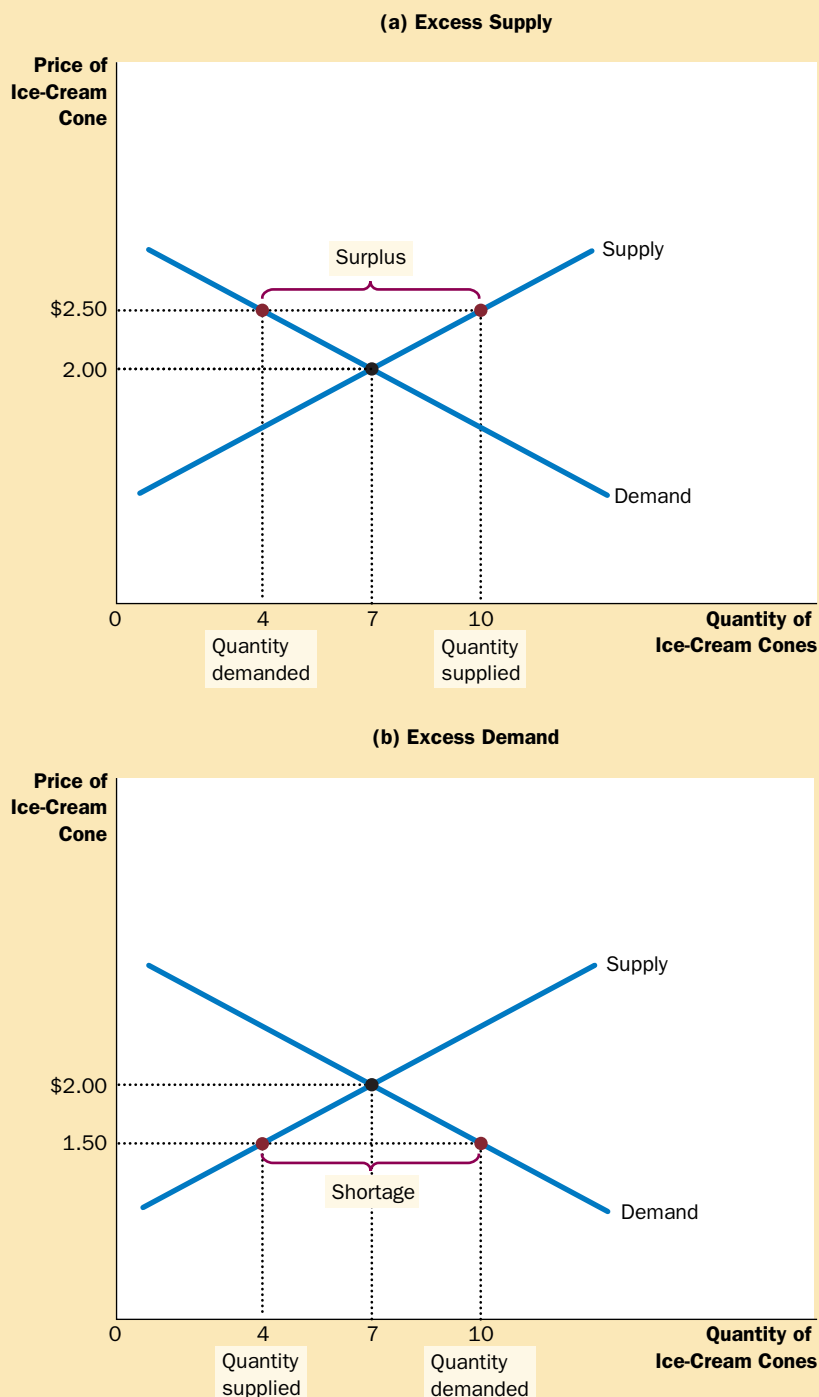
THREE STEPS TO ANALYZING CHANGES IN EQUILIBRIUM

So far we have seen how supply and demand together determine a market's equilibrium, which in turn determines the price of the good and the amount of the good that buyers purchase and sellers produce. Of course, the equilibrium price and quantity depend on the position of the supply and demand curves. When some event shifts one of these curves, the equilibrium in the market changes. The analysis of such a change is called *comparative statics* because it involves comparing two static situations—an old and a new equilibrium.

When analyzing how some event affects a market, we proceed in three steps. First, we decide whether the event shifts the supply curve, the demand curve, or in some cases both curves. Second, we decide whether the curve shifts to the right or to the left. Third, we use the supply-and-demand diagram to examine how the

Figure 4-9

MARKETS NOT IN EQUILIBRIUM. In panel (a), there is a surplus. Because the market price of \$2.50 is above the equilibrium price, the quantity supplied (10 cones) exceeds the quantity demanded (4 cones). Suppliers try to increase sales by cutting the price of a cone, and this moves the price toward its equilibrium level. In panel (b), there is a shortage. Because the market price of \$1.50 is below the equilibrium price, the quantity demanded (10 cones) exceeds the quantity supplied (4 cones). With too many buyers chasing too few goods, suppliers can take advantage of the shortage by raising the price. Hence, in both cases, the price adjustment moves the market toward the equilibrium of supply and demand.





shift affects the equilibrium price and quantity. Table 4-7 summarizes these three steps. To see how this recipe is used, let's consider various events that might affect the market for ice cream.

Example: A Change in Demand Suppose that one summer the weather is very hot. How does this event affect the market for ice cream? To answer this question, let's follow our three steps.

1. The hot weather affects the demand curve by changing people's taste for ice cream. That is, the weather changes the amount of ice cream that people want to buy at any given price. The supply curve is unchanged because the weather does not directly affect the firms that sell ice cream.
2. Because hot weather makes people want to eat more ice cream, the demand curve shifts to the right. Figure 4-10 shows this increase in demand as the shift in the demand curve from D_1 to D_2 . This shift indicates that the quantity of ice cream demanded is higher at every price.
3. As Figure 4-10 shows, the increase in demand raises the equilibrium price from \$2.00 to \$2.50 and the equilibrium quantity from 7 to 10 cones. In other words, the hot weather increases the price of ice cream and the quantity of ice cream sold.

Shifts in Curves versus Movements along Curves Notice that when hot weather drives up the price of ice cream, the quantity of ice cream that firms supply rises, even though the supply curve remains the same. In this case, economists say there has been an increase in "quantity supplied" but no change in "supply."

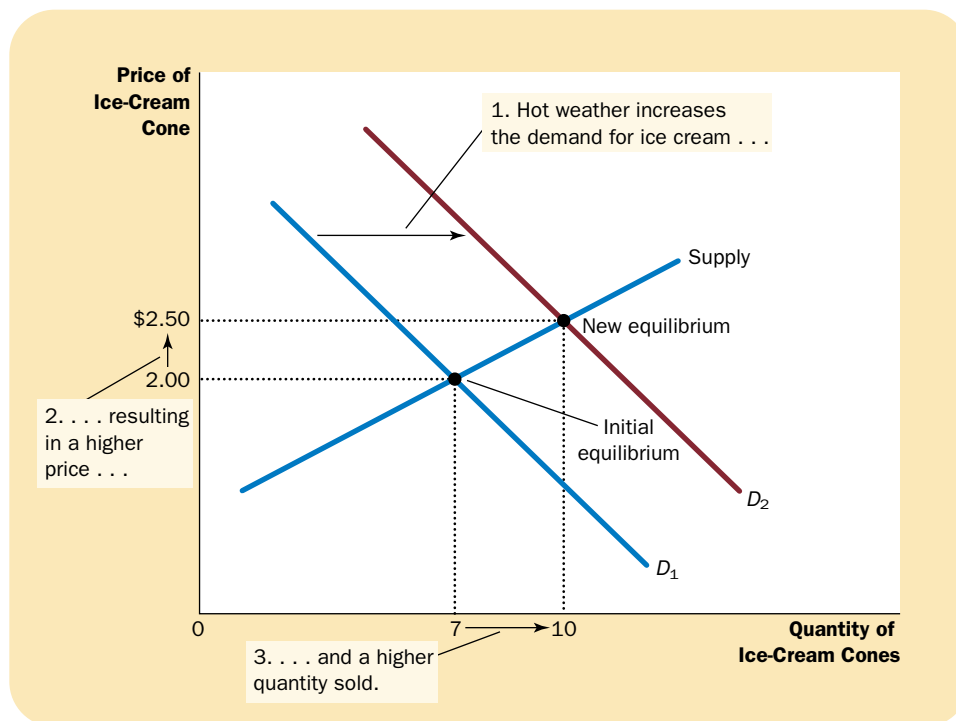
1. Decide whether the event shifts the supply curve or demand curve (or perhaps both).
2. Decide which direction the curve shifts.
3. Use the supply-and-demand diagram to see how the shift changes the equilibrium.

Table 4-7

A THREE-STEP PROGRAM FOR
ANALYZING CHANGES IN
EQUILIBRIUM

Figure 4-10

HOW AN INCREASE IN DEMAND AFFECTS THE EQUILIBRIUM. An event that raises quantity demanded at any given price shifts the demand curve to the right. The equilibrium price and the equilibrium quantity both rise. Here, an abnormally hot summer causes buyers to demand more ice cream. The demand curve shifts from D_1 to D_2 , which causes the equilibrium price to rise from \$2.00 to \$2.50 and the equilibrium quantity to rise from 7 to 10 cones.

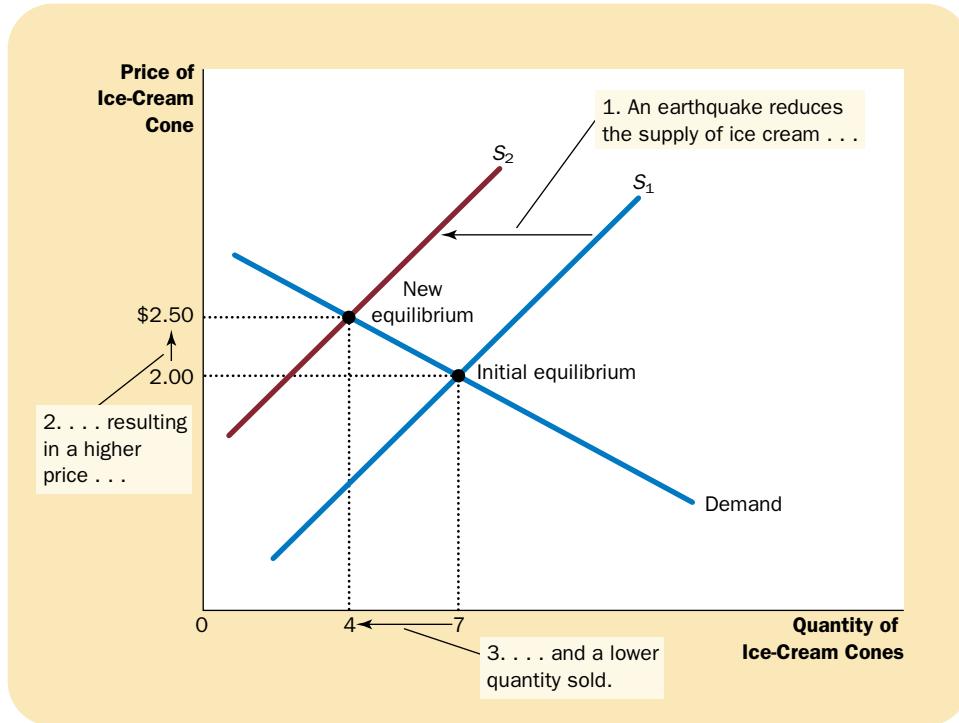


“Supply” refers to the position of the supply curve, whereas the “quantity supplied” refers to the amount suppliers wish to sell. In this example, supply does not change because the weather does not alter firms’ desire to sell at any given price. Instead, the hot weather alters consumers’ desire to buy at any given price and thereby shifts the demand curve. The increase in demand causes the equilibrium price to rise. When the price rises, the quantity supplied rises. This increase in quantity supplied is represented by the movement along the supply curve.

To summarize, a shift *in* the supply curve is called a “change in supply,” and a shift *in* the demand curve is called a “change in demand.” A movement *along* a fixed supply curve is called a “change in the quantity supplied,” and a movement *along* a fixed demand curve is called a “change in the quantity demanded.”

Example: A Change in Supply Suppose that, during another summer, an earthquake destroys several ice-cream factories. How does this event affect the market for ice cream? Once again, to answer this question, we follow our three steps.

1. The earthquake affects the supply curve. By reducing the number of sellers, the earthquake changes the amount of ice cream that firms produce and sell at any given price. The demand curve is unchanged because the earthquake does not directly change the amount of ice cream households wish to buy.
2. The supply curve shifts to the left because, at every price, the total amount that firms are willing and able to sell is reduced. Figure 4-11 illustrates this decrease in supply as a shift in the supply curve from S_1 to S_2 .

**Figure 4-11**

HOW A DECREASE IN SUPPLY AFFECTS THE EQUILIBRIUM.

An event that reduces quantity supplied at any given price shifts the supply curve to the left. The equilibrium price rises, and the equilibrium quantity falls. Here, an earthquake causes sellers to supply less ice cream. The supply curve shifts from S_1 to S_2 , which causes the equilibrium price to rise from \$2.00 to \$2.50 and the equilibrium quantity to fall from 7 to 4 cones.

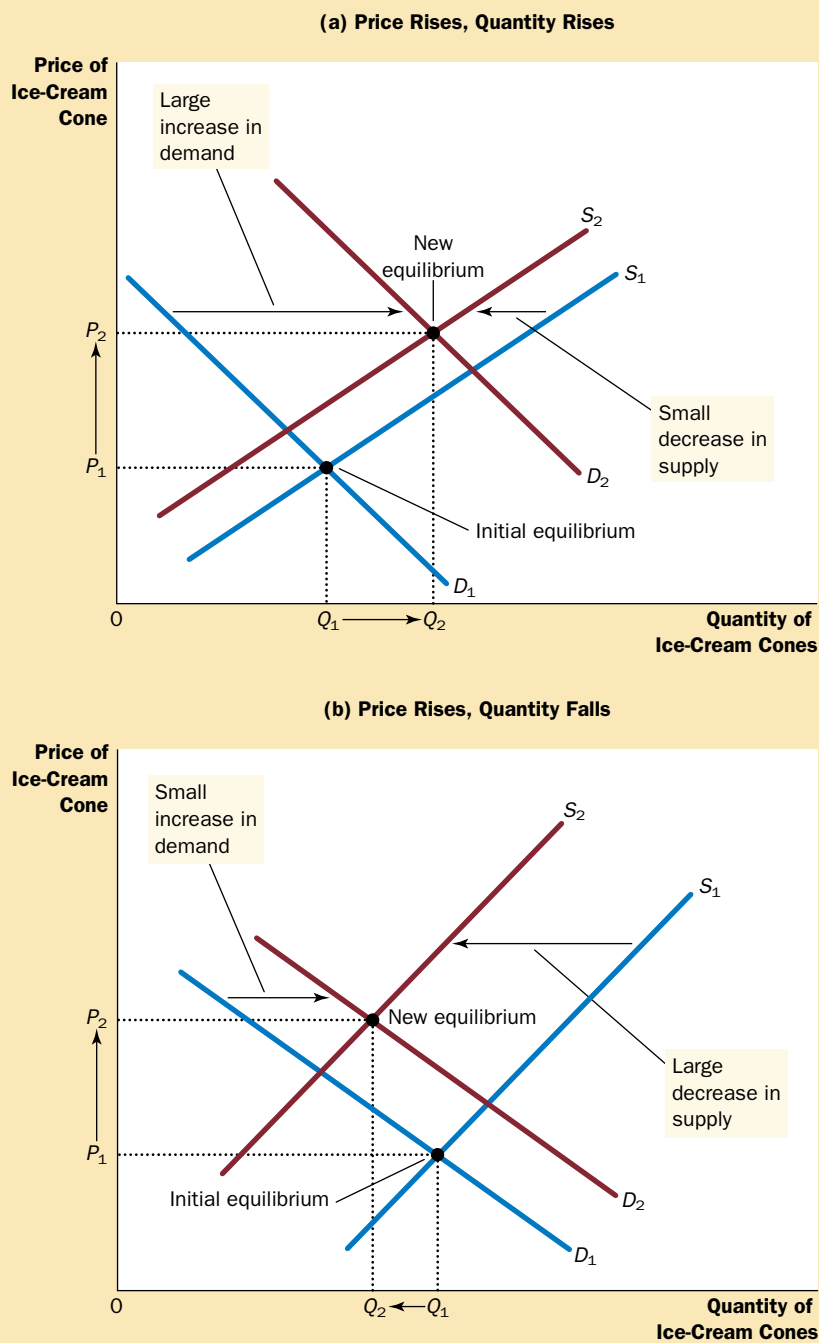
- As Figure 4-11 shows, the shift in the supply curve raises the equilibrium price from \$2.00 to \$2.50 and lowers the equilibrium quantity from 7 to 4 cones. As a result of the earthquake, the price of ice cream rises, and the quantity of ice cream sold falls.

Example: A Change in Both Supply and Demand Now suppose that the hot weather and the earthquake occur at the same time. To analyze this combination of events, we again follow our three steps.

- We determine that both curves must shift. The hot weather affects the demand curve because it alters the amount of ice cream that households want to buy at any given price. At the same time, the earthquake alters the supply curve because it changes the amount of ice cream that firms want to sell at any given price.
- The curves shift in the same directions as they did in our previous analysis: The demand curve shifts to the right, and the supply curve shifts to the left. Figure 4-12 illustrates these shifts.
- As Figure 4-12 shows, there are two possible outcomes that might result, depending on the relative size of the demand and supply shifts. In both cases, the equilibrium price rises. In panel (a), where demand increases substantially while supply falls just a little, the equilibrium quantity also rises. By contrast, in panel (b), where supply falls substantially while demand rises just a little, the equilibrium quantity falls. Thus, these events certainly raise the price of ice cream, but their impact on the amount of ice cream sold is ambiguous.

Figure 4-12

A SHIFT IN BOTH SUPPLY AND DEMAND. Here we observe a simultaneous increase in demand and decrease in supply. Two outcomes are possible. In panel (a), the equilibrium price rises from P_1 to P_2 , and the equilibrium quantity rises from Q_1 to Q_2 . In panel (b), the equilibrium price again rises from P_1 to P_2 , but the equilibrium quantity falls from Q_1 to Q_2 .



Summary We have just seen three examples of how to use supply and demand curves to analyze a change in equilibrium. Whenever an event shifts the supply curve, the demand curve, or perhaps both curves, you can use these tools to predict how the event will alter the amount sold in equilibrium and the price at which the

IN THE NEWS*Mother Nature Shifts
the Supply Curve*

ACCORDING TO OUR ANALYSIS, A NATURAL disaster that reduces supply reduces the quantity sold and raises the price. Here's a recent example.

**4-Day Cold Spell Slams
California: Crops Devastated;
Price of Citrus to Rise**

BY TODD S. PURDUM

A brutal four-day freeze has destroyed more than a third of California's annual

citrus crop, inflicting upwards of a half-billion dollars in damage and raising the prospect of tripled orange prices in supermarkets by next week.

Throughout the Golden State, cold, dry air from the Gulf of Alaska sent temperatures below freezing beginning Monday, with readings in the high teens and low 20's in agriculturally rich Central Valley early today—the worst cold spell since a 10-day freeze in 1990. Farmers frantically ran wind and irrigation machines overnight to keep trees warm, but officials pronounced a near total loss in the valley, and said perhaps half of the state's orange crop was lost as well. . . .

California grows about 80 percent of the nation's oranges eaten as fruit, and 90 percent of lemons, and wholesalers said the retail prices of oranges could triple in the next few days. The price of lemons was certain to rise as well, but the price of orange juice should



be less affected because most juice oranges are grown in Florida.

In some California markets, wholesalers reported that the price of navel oranges had increased to 90 cents a pound on Wednesday from 35 cents on Tuesday.

SOURCE: *The New York Times*, December 25, 1998, p. A1.

Table 4-8

WHAT HAPPENS TO PRICE AND QUANTITY WHEN SUPPLY OR DEMAND SHIFTS?

	NO CHANGE IN SUPPLY	AN INCREASE IN SUPPLY	A DECREASE IN SUPPLY
NO CHANGE IN DEMAND	P same Q same	P down Q up	P up Q down
AN INCREASE IN DEMAND	P up Q up	P ambiguous Q up	P up Q ambiguous
A DECREASE IN DEMAND	P down Q down	P down Q ambiguous	P ambiguous Q down

good is sold. Table 4-8 shows the predicted outcome for any combination of shifts in the two curves. To make sure you understand how to use the tools of supply and demand, pick a few entries in this table and make sure you can explain to yourself why the table contains the prediction it does.

QUICK QUIZ: Analyze what happens to the market for pizza if the price of tomatoes rises. ♦ Analyze what happens to the market for pizza if the price of hamburgers falls.

CONCLUSION: HOW PRICES ALLOCATE RESOURCES

This chapter has analyzed supply and demand in a single market. Although our discussion has centered around the market for ice cream, the lessons learned here apply in most other markets as well. Whenever you go to a store to buy something, you are contributing to the demand for that item. Whenever you look for a job, you are contributing to the supply of labor services. Because supply and demand are such pervasive economic phenomena, the model of supply and demand is a powerful tool for analysis. We will be using this model repeatedly in the following chapters.



One of the *Ten Principles of Economics* discussed in Chapter 1 is that markets are usually a good way to organize economic activity. Although it is still too early to judge whether market outcomes are good or bad, in this chapter we have begun to see how markets work. In any economic system, scarce resources have to be allocated among competing uses. Market economies harness the forces of supply and demand to serve that end. Supply and demand together determine the prices of the economy's many different goods and services; prices in turn are the signals that guide the allocation of resources.

For example, consider the allocation of beachfront land. Because the amount of this land is limited, not everyone can enjoy the luxury of living by the beach. Who gets this resource? The answer is: whoever is willing and able to pay the price. The price of beachfront land adjusts until the quantity of land demanded exactly balances the quantity supplied. Thus, in market economies, prices are the mechanism for rationing scarce resources.

Similarly, prices determine who produces each good and how much is produced. For instance, consider farming. Because we need food to survive, it is crucial that some people work on farms. What determines who is a farmer and who is not? In a free society, there is no government planning agency making this decision and ensuring an adequate supply of food. Instead, the allocation of workers to farms is based on the job decisions of millions of workers. This decentralized system works well because these decisions depend on prices. The prices of food and the wages of farmworkers (the price of their labor) adjust to ensure that enough people choose to be farmers.

If a person had never seen a market economy in action, the whole idea might seem preposterous. Economies are large groups of people engaged in many interdependent activities. What prevents decentralized decisionmaking from degenerating into chaos? What coordinates the actions of the millions of people with their varying abilities and desires? What ensures that what needs to get done does in fact get done? The answer, in a word, is *prices*. If market economies are guided by an invisible hand, as Adam Smith famously suggested, then the price system is the baton that the invisible hand uses to conduct the economic orchestra.



"Two dollars."



"—and seventy-five cents."

Summary

- ◆ Economists use the model of supply and demand to analyze competitive markets. In a competitive market, there are many buyers and sellers, each of whom has little or no influence on the market price.
- ◆ The demand curve shows how the quantity of a good demanded depends on the price. According to the law of demand, as the price of a good falls, the quantity demanded rises. Therefore, the demand curve slopes downward.
- ◆ In addition to price, other determinants of the quantity demanded include income, tastes, expectations, and the prices of substitutes and complements. If one of these other determinants changes, the demand curve shifts.
- ◆ The supply curve shows how the quantity of a good supplied depends on the price. According to the law of supply, as the price of a good rises, the quantity supplied rises. Therefore, the supply curve slopes upward.
- ◆ In addition to price, other determinants of the quantity supplied include input prices, technology, and expectations. If one of these other determinants changes, the supply curve shifts.
- ◆ The intersection of the supply and demand curves determines the market equilibrium. At the equilibrium price, the quantity demanded equals the quantity supplied.
- ◆ The behavior of buyers and sellers naturally drives markets toward their equilibrium. When the market price is above the equilibrium price, there is a surplus of the good, which causes the market price to fall. When the market price is below the equilibrium price, there is a shortage, which causes the market price to rise.
- ◆ To analyze how any event influences a market, we use the supply-and-demand diagram to examine how the event affects the equilibrium price and quantity. To do this we follow three steps. First, we decide whether the event shifts the supply curve or the demand curve (or both). Second, we decide which direction the curve shifts. Third, we compare the new equilibrium with the old equilibrium.
- ◆ In market economies, prices are the signals that guide economic decisions and thereby allocate scarce resources. For every good in the economy, the price ensures that supply and demand are in balance. The equilibrium price then determines how much of the good buyers choose to purchase and how much sellers choose to produce.

Key Concepts

market, p. 66	complements, p. 68	supply curve, p. 76
competitive market, p. 66	demand schedule, p. 69	equilibrium, p. 80
quantity demanded, p. 67	demand curve, p. 70	equilibrium price, p. 80
law of demand, p. 68	<i>ceteris paribus</i> , p. 70	equilibrium quantity, p. 80
normal good, p. 68	quantity supplied, p. 75	surplus, p. 81
inferior good, p. 68	law of supply, p. 75	shortage, p. 81
substitutes, p. 68	supply schedule, p. 76	law of supply and demand, p. 81

Questions for Review

1. What is a competitive market? Briefly describe the types of markets other than perfectly competitive markets.
2. What determines the quantity of a good that buyers demand?
3. What are the demand schedule and the demand curve, and how are they related? Why does the demand curve slope downward?
4. Does a change in consumers' tastes lead to a movement along the demand curve or a shift in the demand curve? Does a change in price lead to a movement along the demand curve or a shift in the demand curve?
5. Popeye's income declines and, as a result, he buys more spinach. Is spinach an inferior or a normal good? What happens to Popeye's demand curve for spinach?
6. What determines the quantity of a good that sellers supply?
7. What are the supply schedule and the supply curve, and how are they related? Why does the supply curve slope upward?
8. Does a change in producers' technology lead to a movement along the supply curve or a shift in the supply curve? Does a change in price lead to a movement along the supply curve or a shift in the supply curve?
9. Define the equilibrium of a market. Describe the forces that move a market toward its equilibrium.
10. Beer and pizza are complements because they are often enjoyed together. When the price of beer rises, what happens to the supply, demand, quantity supplied, quantity demanded, and the price in the market for pizza?
11. Describe the role of prices in market economies.

Problems and Applications

1. Explain each of the following statements using supply-and-demand diagrams.
 - a. When a cold snap hits Florida, the price of orange juice rises in supermarkets throughout the country.
 - b. When the weather turns warm in New England every summer, the prices of hotel rooms in Caribbean resorts plummet.
 - c. When a war breaks out in the Middle East, the price of gasoline rises, while the price of a used Cadillac falls.
2. "An increase in the demand for notebooks raises the quantity of notebooks demanded, but not the quantity supplied." Is this statement true or false? Explain.
3. Consider the market for minivans. For each of the events listed here, identify which of the determinants of demand or supply are affected. Also indicate whether demand or supply is increased or decreased. Then show the effect on the price and quantity of minivans.
 - a. People decide to have more children.

- b. A strike by steelworkers raises steel prices.
 - c. Engineers develop new automated machinery for the production of minivans.
 - d. The price of station wagons rises.
 - e. A stock-market crash lowers people's wealth.
4. During the 1990s, technological advance reduced the cost of computer chips. How do you think this affected the market for computers? For computer software? For typewriters?
 5. Using supply-and-demand diagrams, show the effect of the following events on the market for sweatshirts.
 - a. A hurricane in South Carolina damages the cotton crop.
 - b. The price of leather jackets falls.
 - c. All colleges require morning calisthenics in appropriate attire.
 - d. New knitting machines are invented.
 6. Suppose that in the year 2005 the number of births is temporarily high. How does this baby boom affect the price of baby-sitting services in 2010 and 2020? (Hint: 5-year-olds need baby-sitters, whereas 15-year-olds can be baby-sitters.)
 7. Ketchup is a complement (as well as a condiment) for hot dogs. If the price of hot dogs rises, what happens to the market for ketchup? For tomatoes? For tomato juice? For orange juice?
 8. The case study presented in the chapter discussed cigarette taxes as a way to reduce smoking. Now think about the markets for other tobacco products such as cigars and chewing tobacco.
 - a. Are these goods substitutes or complements for cigarettes?
 - b. Using a supply-and-demand diagram, show what happens in the markets for cigars and chewing tobacco if the tax on cigarettes is increased.
 - c. If policymakers wanted to reduce total tobacco consumption, what policies could they combine with the cigarette tax?
 9. The market for pizza has the following demand and supply schedules:

PRICE	QUANTITY DEMANDED	QUANTITY SUPPLIED
\$4	135	26
5	104	53
6	81	81
7	68	98
8	53	110
9	39	121

Graph the demand and supply curves. What is the equilibrium price and quantity in this market? If the actual price in this market were *above* the equilibrium price, what would drive the market toward the equilibrium? If the actual price in this market were *below* the equilibrium price, what would drive the market toward the equilibrium?

10. Because bagels and cream cheese are often eaten together, they are complements.
 - a. We observe that both the equilibrium price of cream cheese and the equilibrium quantity of bagels have risen. What could be responsible for this pattern—a fall in the price of flour or a fall in the price of milk? Illustrate and explain your answer.
 - b. Suppose instead that the equilibrium price of cream cheese has risen but the equilibrium quantity of bagels has fallen. What could be responsible for this pattern—a rise in the price of flour or a rise in the price of milk? Illustrate and explain your answer.
11. Suppose that the price of basketball tickets at your college is determined by market forces. Currently, the demand and supply schedules are as follows:

PRICE	QUANTITY DEMANDED	QUANTITY SUPPLIED
\$ 4	10,000	8,000
8	8,000	8,000
12	6,000	8,000
16	4,000	8,000
20	2,000	8,000

- a. Draw the demand and supply curves. What is unusual about this supply curve? Why might this be true?
- b. What are the equilibrium price and quantity of tickets?
- c. Your college plans to increase total enrollment next year by 5,000 students. The additional students will have the following demand schedule:

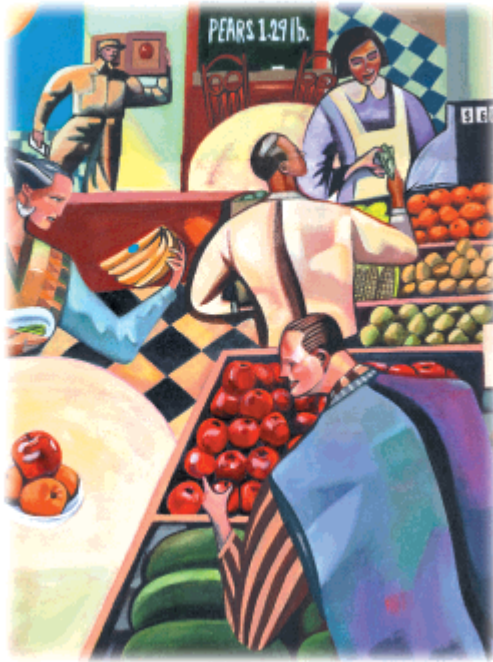
PRICE	QUANTITY DEMANDED
\$ 4	4,000
8	3,000
12	2,000
16	1,000
20	0

Now add the old demand schedule and the demand schedule for the new students to calculate the new demand schedule for the entire college. What will be the new equilibrium price and quantity?

12. An article in *The New York Times* described a successful marketing campaign by the French champagne industry.

The article noted that “many executives felt giddy about the stratospheric champagne prices. But they also feared that such sharp price increases would cause demand to decline, which would then cause prices to plunge.” What mistake are the executives making in their analysis of the situation? Illustrate your answer with a graph.

5



ELASTICITY AND ITS APPLICATION

Imagine yourself as a Kansas wheat farmer. Because you earn all your income from selling wheat, you devote much effort to making your land as productive as it can be. You monitor weather and soil conditions, check your fields for pests and disease, and study the latest advances in farm technology. You know that the more wheat you grow, the more you will have to sell after the harvest, and the higher will be your income and your standard of living.

One day Kansas State University announces a major discovery. Researchers in its agronomy department have devised a new hybrid of wheat that raises the amount farmers can produce from each acre of land by 20 percent. How should you react to this news? Should you use the new hybrid? Does this discovery make you better off or worse off than you were before? In this chapter we will see that these questions can have surprising answers. The surprise will come from

**IN THIS CHAPTER
YOU WILL . . .**

***Learn the meaning
of the elasticity of
demand***

***Examine what
determines the
elasticity of demand***

***Learn the meaning
of the elasticity of
supply***

***Examine what
determines the
elasticity of supply***

***Apply the concept of
elasticity in three
very different
markets***

applying the most basic tools of economics—supply and demand—to the market for wheat.

The previous chapter introduced supply and demand. In any competitive market, such as the market for wheat, the upward-sloping supply curve represents the behavior of sellers, and the downward-sloping demand curve represents the behavior of buyers. The price of the good adjusts to bring the quantity supplied and quantity demanded of the good into balance. To apply this basic analysis to understand the impact of the agronomists' discovery, we must first develop one more tool: the concept of *elasticity*. Elasticity, a measure of how much buyers and sellers respond to changes in market conditions, allows us to analyze supply and demand with greater precision.

THE ELASTICITY OF DEMAND

When we discussed the determinants of demand in Chapter 4, we noted that buyers usually demand more of a good when its price is lower, when their incomes are higher, when the prices of substitutes for the good are higher, or when the prices of complements of the good are lower. Our discussion of demand was qualitative, not quantitative. That is, we discussed the direction in which the quantity demanded moves, but not the size of the change. To measure how much demand responds to changes in its determinants, economists use the concept of **elasticity**.

elasticity

a measure of the responsiveness of quantity demanded or quantity supplied to one of its determinants

THE PRICE ELASTICITY OF DEMAND AND ITS DETERMINANTS

The law of demand states that a fall in the price of a good raises the quantity demanded. The **price elasticity of demand** measures how much the quantity demanded responds to a change in price. Demand for a good is said to be *elastic* if the quantity demanded responds substantially to changes in the price. Demand is said to be *inelastic* if the quantity demanded responds only slightly to changes in the price.

What determines whether the demand for a good is elastic or inelastic? Because the demand for any good depends on consumer preferences, the price elasticity of demand depends on the many economic, social, and psychological forces that shape individual desires. Based on experience, however, we can state some general rules about what determines the price elasticity of demand.

Necessities versus Luxuries Necessities tend to have inelastic demands, whereas luxuries have elastic demands. When the price of a visit to the doctor rises, people will not dramatically alter the number of times they go to the doctor, although they might go somewhat less often. By contrast, when the price of sailboats rises, the quantity of sailboats demanded falls substantially. The reason is that most people view doctor visits as a necessity and sailboats as a luxury. Of course, whether a good is a necessity or a luxury depends not on the intrinsic properties of the good but on the preferences of the buyer. For an avid sailor with

price elasticity of demand

a measure of how much the quantity demanded of a good responds to a change in the price of that good, computed as the percentage change in quantity demanded divided by the percentage change in price

little concern over his health, sailboats might be a necessity with inelastic demand and doctor visits a luxury with elastic demand.

Availability of Close Substitutes Goods with close substitutes tend to have more elastic demand because it is easier for consumers to switch from that good to others. For example, butter and margarine are easily substitutable. A small increase in the price of butter, assuming the price of margarine is held fixed, causes the quantity of butter sold to fall by a large amount. By contrast, because eggs are a food without a close substitute, the demand for eggs is probably less elastic than the demand for butter.

Definition of the Market The elasticity of demand in any market depends on how we draw the boundaries of the market. Narrowly defined markets tend to have more elastic demand than broadly defined markets, because it is easier to find close substitutes for narrowly defined goods. For example, food, a broad category, has a fairly inelastic demand because there are no good substitutes for food. Ice cream, a more narrow category, has a more elastic demand because it is easy to substitute other desserts for ice cream. Vanilla ice cream, a very narrow category, has a very elastic demand because other flavors of ice cream are almost perfect substitutes for vanilla.

Time Horizon Goods tend to have more elastic demand over longer time horizons. When the price of gasoline rises, the quantity of gasoline demanded falls only slightly in the first few months. Over time, however, people buy more fuel-efficient cars, switch to public transportation, and move closer to where they work. Within several years, the quantity of gasoline demanded falls substantially.

COMPUTING THE PRICE ELASTICITY OF DEMAND

Now that we have discussed the price elasticity of demand in general terms, let's be more precise about how it is measured. Economists compute the price elasticity of demand as the percentage change in the quantity demanded divided by the percentage change in the price. That is,

$$\text{Price elasticity of demand} = \frac{\text{Percentage change in quantity demanded}}{\text{Percentage change in price}}.$$

For example, suppose that a 10-percent increase in the price of an ice-cream cone causes the amount of ice cream you buy to fall by 20 percent. We calculate your elasticity of demand as

$$\text{Price elasticity of demand} = \frac{20 \text{ percent}}{10 \text{ percent}} = 2.$$

In this example, the elasticity is 2, reflecting that the change in the quantity demanded is proportionately twice as large as the change in the price.

Because the quantity demanded of a good is negatively related to its price, the percentage change in quantity will always have the opposite sign as the

percentage change in price. In this example, the percentage change in price is a *positive* 10 percent (reflecting an increase), and the percentage change in quantity demanded is a *negative* 20 percent (reflecting a decrease). For this reason, price elasticities of demand are sometimes reported as negative numbers. In this book we follow the common practice of dropping the minus sign and reporting all price elasticities as positive numbers. (Mathematicians call this the *absolute value*.) With this convention, a larger price elasticity implies a greater responsiveness of quantity demanded to price.

THE MIDPOINT METHOD: A BETTER WAY TO CALCULATE PERCENTAGE CHANGES AND ELASTICITIES

If you try calculating the price elasticity of demand between two points on a demand curve, you will quickly notice an annoying problem: The elasticity from point A to point B seems different from the elasticity from point B to point A. For example, consider these numbers:

Point A:	Price = \$4	Quantity = 120
Point B:	Price = \$6	Quantity = 80

Going from point A to point B, the price rises by 50 percent, and the quantity falls by 33 percent, indicating that the price elasticity of demand is $33/50$, or 0.66. By contrast, going from point B to point A, the price falls by 33 percent, and the quantity rises by 50 percent, indicating that the price elasticity of demand is $50/33$, or 1.5.

One way to avoid this problem is to use the *midpoint method* for calculating elasticities. Rather than computing a percentage change using the standard way (by dividing the change by the initial level), the midpoint method computes a percentage change by dividing the change by the midpoint of the initial and final levels. For instance, \$5 is the midpoint of \$4 and \$6. Therefore, according to the midpoint method, a change from \$4 to \$6 is considered a 40 percent rise, because $(6 - 4)/5 \times 100 = 40$. Similarly, a change from \$6 to \$4 is considered a 40 percent fall.

Because the midpoint method gives the same answer regardless of the direction of change, it is often used when calculating the price elasticity of demand between two points. In our example, the midpoint between point A and point B is:

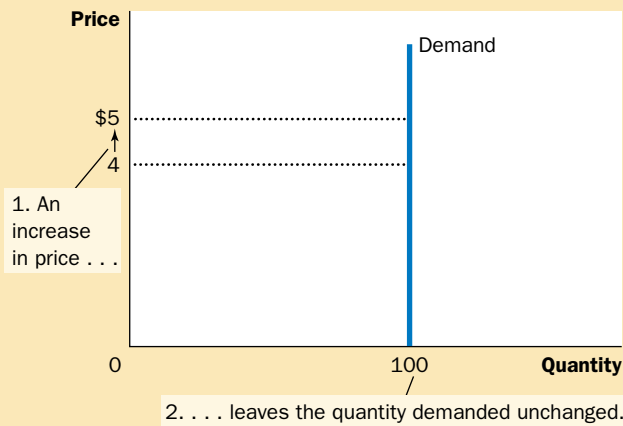
Midpoint:	Price = \$5	Quantity = 100

According to the midpoint method, when going from point A to point B, the price rises by 40 percent, and the quantity falls by 40 percent. Similarly, when going from point B to point A, the price falls by 40 percent, and the quantity rises by 40 percent. In both directions, the price elasticity of demand equals 1.

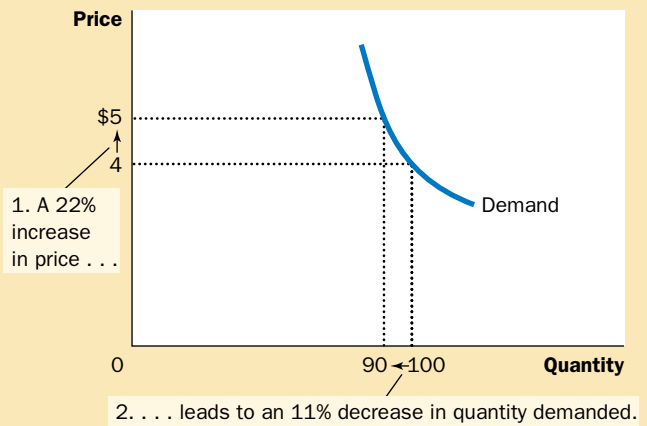
We can express the midpoint method with the following formula for the price elasticity of demand between two points, denoted (Q_1, P_1) and (Q_2, P_2) :

$$\text{Price elasticity of demand} = \frac{(Q_2 - Q_1)/[(Q_2 + Q_1)/2]}{(P_2 - P_1)/[(P_2 + P_1)/2]}.$$

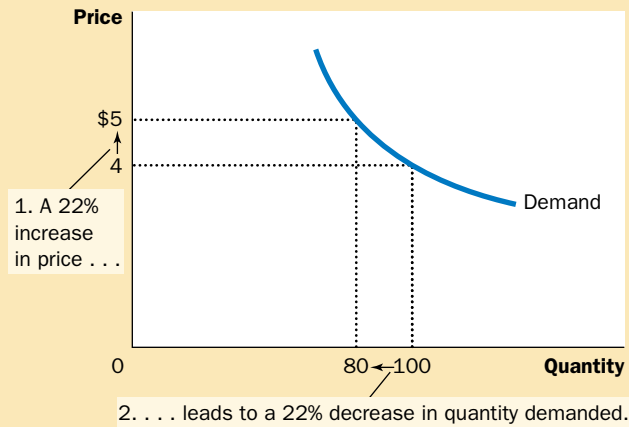
(a) Perfectly Inelastic Demand: Elasticity Equals 0



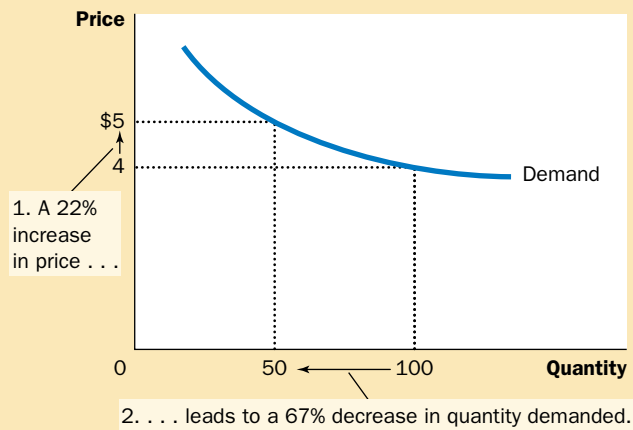
(b) Inelastic Demand: Elasticity Is Less Than 1



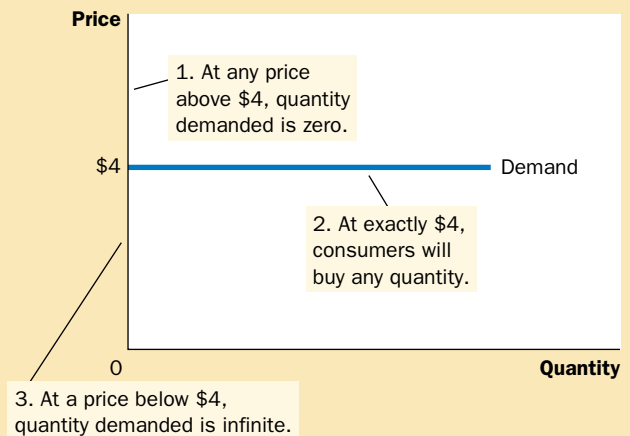
(c) Unit Elastic Demand: Elasticity Equals 1



(d) Elastic Demand: Elasticity Is Greater Than 1



(e) Perfectly Elastic Demand: Elasticity Equals Infinity



THE PRICE ELASTICITY OF DEMAND. The price elasticity of demand determines whether the demand curve is steep or flat. Note that all percentage changes are calculated using the midpoint method.

Figure 5-1

The numerator is the percentage change in quantity computed using the midpoint method, and the denominator is the percentage change in price computed using the midpoint method. If you ever need to calculate elasticities, you should use this formula.

Throughout this book, however, we only rarely need to perform such calculations. For our purposes, what elasticity represents—the responsiveness of quantity demanded to price—is more important than how it is calculated.

THE VARIETY OF DEMAND CURVES

Economists classify demand curves according to their elasticity. Demand is *elastic* when the elasticity is greater than 1, so that quantity moves proportionately more than the price. Demand is *inelastic* when the elasticity is less than 1, so that quantity moves proportionately less than the price. If the elasticity is exactly 1, so that quantity moves the same amount proportionately as price, demand is said to have *unit elasticity*.

Because the price elasticity of demand measures how much quantity demanded responds to changes in the price, it is closely related to the slope of the demand curve. The following rule of thumb is a useful guide: The flatter is the demand curve that passes through a given point, the greater is the price elasticity of demand. The steeper is the demand curve that passes through a given point, the smaller is the price elasticity of demand.

Figure 5-1 shows five cases. In the extreme case of a zero elasticity, demand is *perfectly inelastic*, and the demand curve is vertical. In this case, regardless of the price, the quantity demanded stays the same. As the elasticity rises, the demand curve gets flatter and flatter. At the opposite extreme, demand is *perfectly elastic*. This occurs as the price elasticity of demand approaches infinity and the demand curve becomes horizontal, reflecting the fact that very small changes in the price lead to huge changes in the quantity demanded.

Finally, if you have trouble keeping straight the terms *elastic* and *inelastic*, here's a memory trick for you: *I*nelastic curves, such as in panel (a) of Figure 5-1, look like the letter *I*. *E*lastic curves, as in panel (e), look like the letter *E*. This is not a deep insight, but it might help on your next exam.

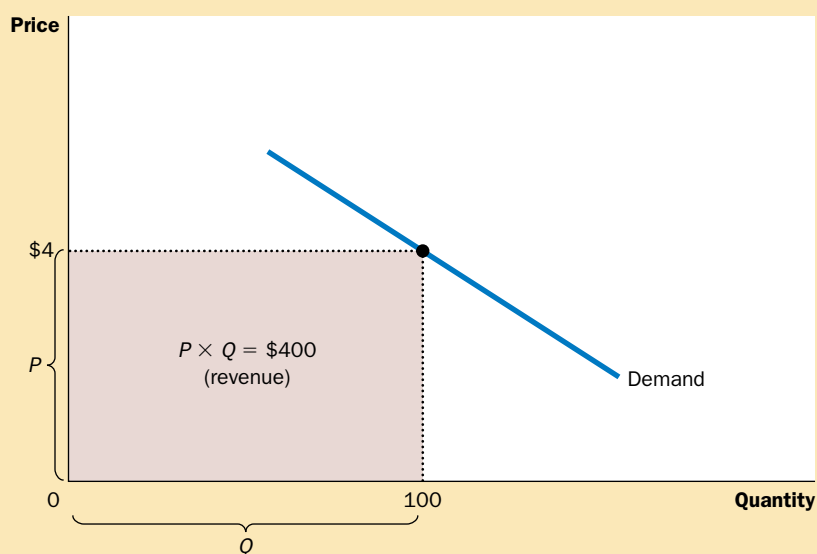
TOTAL REVENUE AND THE PRICE ELASTICITY OF DEMAND

total revenue

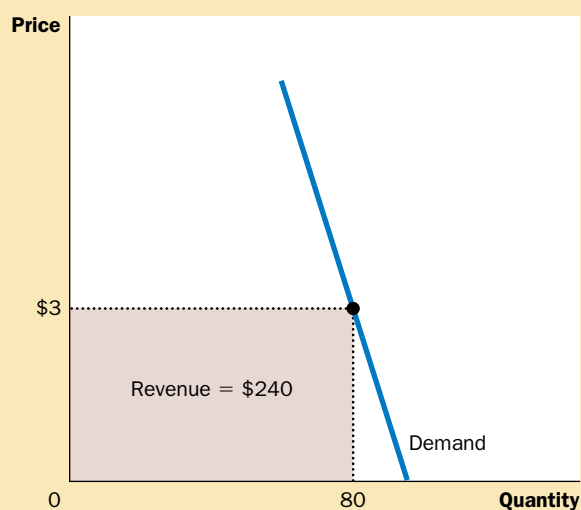
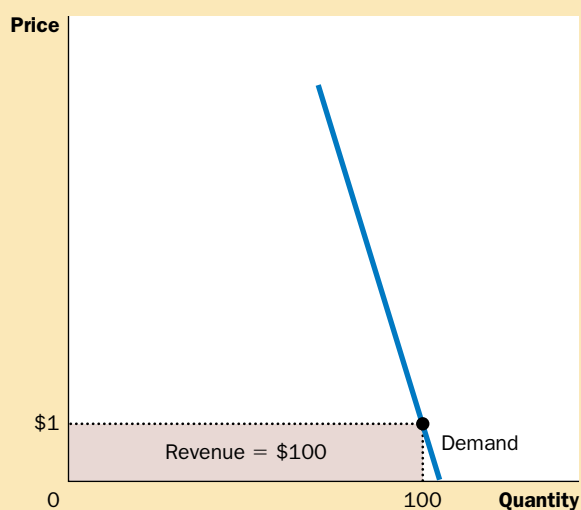
the amount paid by buyers and received by sellers of a good, computed as the price of the good times the quantity sold

When studying changes in supply or demand in a market, one variable we often want to study is **total revenue**, the amount paid by buyers and received by sellers of the good. In any market, total revenue is $P \times Q$, the price of the good times the quantity of the good sold. We can show total revenue graphically, as in Figure 5-2. The height of the box under the demand curve is P , and the width is Q . The area of this box, $P \times Q$, equals the total revenue in this market. In Figure 5-2, where $P = \$4$ and $Q = 100$, total revenue is $\$4 \times 100$, or \$400.

How does total revenue change as one moves along the demand curve? The answer depends on the price elasticity of demand. If demand is inelastic, as in Figure 5-3, then an increase in the price causes an increase in total revenue. Here an increase in price from \$1 to \$3 causes the quantity demanded to fall only from 100

Figure 5-2

TOTAL REVENUE. The total amount paid by buyers, and received as revenue by sellers, equals the area of the box under the demand curve, $P \times Q$. Here, at a price of \$4, the quantity demanded is 100, and total revenue is \$400.



HOW TOTAL REVENUE CHANGES WHEN PRICE CHANGES: INELASTIC DEMAND. With an inelastic demand curve, an increase in the price leads to a decrease in quantity demanded that is proportionately smaller. Therefore, total revenue (the product of price and quantity) increases. Here, an increase in the price from \$1 to \$3 causes the quantity demanded to fall from 100 to 80, and total revenue rises from \$100 to \$240.

Figure 5-3

to 80, and so total revenue rises from \$100 to \$240. An increase in price raises $P \times Q$ because the fall in Q is proportionately smaller than the rise in P .

We obtain the opposite result if demand is elastic: An increase in the price causes a decrease in total revenue. In Figure 5-4, for instance, when the price rises from \$4 to \$5, the quantity demanded falls from 50 to 20, and so total revenue falls from \$200 to \$100. Because demand is elastic, the reduction in the quantity demanded is so great that it more than offsets the increase in the price. That is, an increase in price reduces $P \times Q$ because the fall in Q is proportionately greater than the rise in P .

Although the examples in these two figures are extreme, they illustrate a general rule:

- ◆ When a demand curve is inelastic (a price elasticity less than 1), a price increase raises total revenue, and a price decrease reduces total revenue.
- ◆ When a demand curve is elastic (a price elasticity greater than 1), a price increase reduces total revenue, and a price decrease raises total revenue.
- ◆ In the special case of unit elastic demand (a price elasticity exactly equal to 1), a change in the price does not affect total revenue.

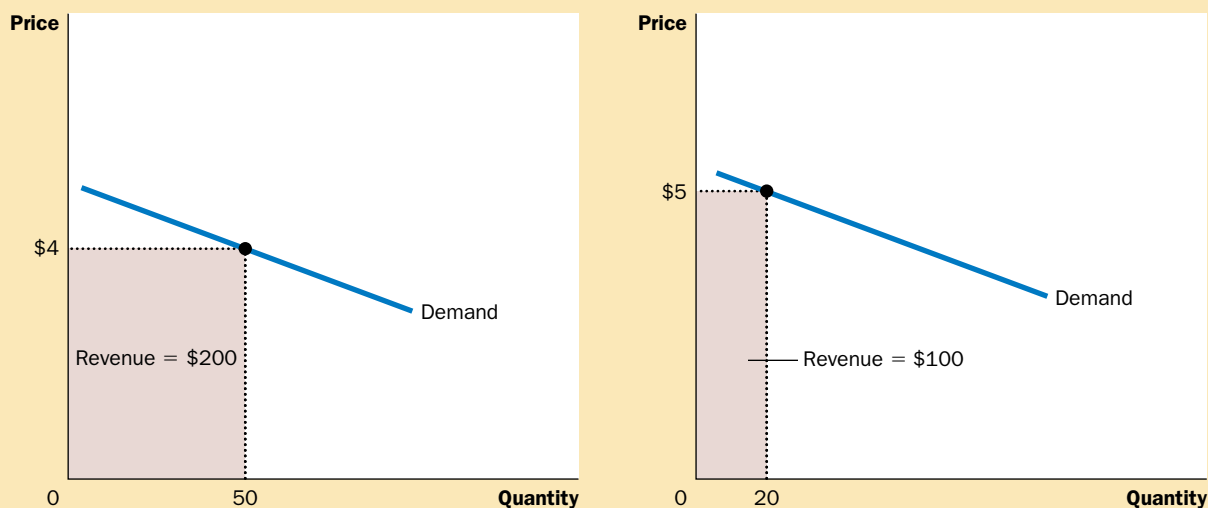
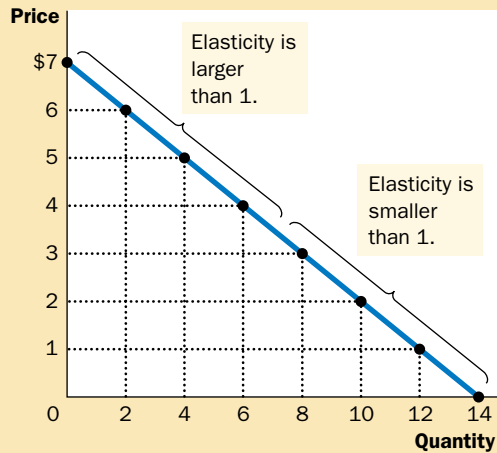


Figure 5-4

HOW TOTAL REVENUE CHANGES WHEN PRICE CHANGES: ELASTIC DEMAND. With an elastic demand curve, an increase in the price leads to a decrease in quantity demanded that is proportionately larger. Therefore, total revenue (the product of price and quantity) decreases. Here, an increase in the price from \$4 to \$5 causes the quantity demanded to fall from 50 to 20, so total revenue falls from \$200 to \$100.

**Figure 5-5**

A LINEAR DEMAND CURVE.
The slope of a linear demand curve is constant, but its elasticity is not.

PRICE	QUANTITY	TOTAL REVENUE (PRICE × QUANTITY)	PERCENT CHANGE IN PRICE	PERCENT CHANGE IN QUANTITY	ELASTICITY	DESCRIPTION
\$7	0	\$ 0	15	200	13.0	Elastic
6	2	12	18	67	3.7	Elastic
5	4	20	22	40	1.8	Elastic
4	6	24	29	29	1.0	Unit elastic
3	8	24	40	22	0.6	Inelastic
2	10	20	67	18	0.3	Inelastic
1	12	12	200	15	0.1	Inelastic
0	14	0				

COMPUTING THE ELASTICITY OF A LINEAR DEMAND CURVE

Table 5-1

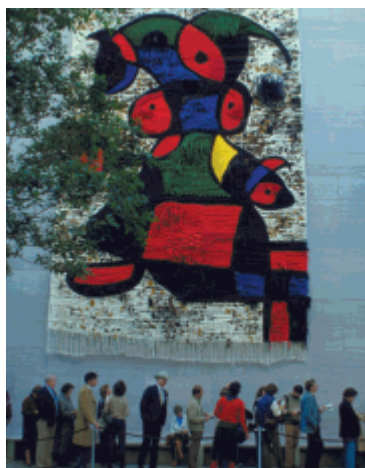
NOTE: Elasticity is calculated here using the midpoint method.

ELASTICITY AND TOTAL REVENUE ALONG A LINEAR DEMAND CURVE

Although some demand curves have an elasticity that is the same along the entire curve, that is not always the case. An example of a demand curve along which elasticity changes is a straight line, as shown in Figure 5-5. A linear demand curve has a constant slope. Recall that slope is defined as “rise over run,” which here is the ratio of the change in price (“rise”) to the change in quantity (“run”). This particular demand curve’s slope is constant because each \$1 increase in price causes the same 2-unit decrease in the quantity demanded.

Even though the slope of a linear demand curve is constant, the elasticity is not. The reason is that the slope is the ratio of *changes* in the two variables, whereas the elasticity is the ratio of *percentage changes* in the two variables. You can see this most easily by looking at Table 5-1. This table shows the demand schedule for the linear demand curve in Figure 5-5 and calculates the price elasticity of demand using the midpoint method discussed earlier. At points with a low price and high quantity, the demand curve is inelastic. At points with a high price and low quantity, the demand curve is elastic.

Table 5-1 also presents total revenue at each point on the demand curve. These numbers illustrate the relationship between total revenue and elasticity. When the price is \$1, for instance, demand is inelastic, and a price increase to \$2 raises total revenue. When the price is \$5, demand is elastic, and a price increase to \$6 reduces total revenue. Between \$3 and \$4, demand is exactly unit elastic, and total revenue is the same at these two prices.



IF THE PRICE OF ADMISSION WERE HIGHER, HOW MUCH SHORTER WOULD THIS LINE BECOME?

CASE STUDY PRICING ADMISSION TO A MUSEUM

You are curator of a major art museum. Your director of finance tells you that the museum is running short of funds and suggests that you consider changing the price of admission to increase total revenue. What do you do? Do you raise the price of admission, or do you lower it?

The answer depends on the elasticity of demand. If the demand for visits to the museum is inelastic, then an increase in the price of admission would increase total revenue. But if the demand is elastic, then an increase in price would cause the number of visitors to fall by so much that total revenue would decrease. In this case, you should cut the price. The number of visitors would rise by so much that total revenue would increase.

To estimate the price elasticity of demand, you would need to turn to your statisticians. They might use historical data to study how museum attendance varied from year to year as the admission price changed. Or they might use data on attendance at the various museums around the country to see how the admission price affects attendance. In studying either of these sets of data, the statisticians would need to take account of other factors that affect attendance—weather, population, size of collection, and so forth—to isolate the effect of price. In the end, such data analysis would provide an estimate of the price elasticity of demand, which you could use in deciding how to respond to your financial problem.

OTHER DEMAND ELASTICITIES

In addition to the price elasticity of demand, economists also use other elasticities to describe the behavior of buyers in a market.

Income elasticity of demand

a measure of how much the quantity demanded of a good responds to a change in consumers' income, computed as the percentage change in quantity demanded divided by the percentage change in income

The Income Elasticity of Demand Economists use the **income elasticity of demand** to measure how the quantity demanded changes as consumer income changes. The income elasticity is the percentage change in quantity demanded divided by the percentage change in income. That is,

IN THE NEWS*On the Road
with Elasticity*

HOW SHOULD A FIRM THAT OPERATES A private toll road set a price for its service? As the following article makes clear, answering this question requires an understanding of the demand curve and its elasticity.

**For Whom the Booth Tolls,
Price Really Does Matter**

BY STEVEN PEARLSTEIN

All businesses face a similar question: What price for their product will generate the maximum profit?

The answer is not always obvious: Raising the price of something often has the effect of reducing sales as price-sensitive consumers seek alternatives or simply do without. For every product, the extent of that sensitivity is different. The trick is to find the point for each where the ideal tradeoff between profit margin and sales volume is achieved.

Right now, the developers of a new private toll road between Leesburg and

Washington-Dulles International Airport are trying to discern the magic point. The group originally projected that it could charge nearly \$2 for the 14-mile one-way trip, while attracting 34,000 trips on an average day from overcrowded public roads such as nearby Route 7. But after spending \$350 million to build their much heralded "Greenway," they discovered to their dismay that only about a third that number of commuters were willing to pay that much to shave 20 minutes off their daily commute. . . .

It was only when the company, in desperation, lowered the toll to \$1 that it came even close to attracting the expected traffic flows.

Although the Greenway still is losing money, it is clearly better off at this new point on the demand curve than it was when it first opened. Average daily revenue today is \$22,000, compared with \$14,875 when the "special introductory" price was \$1.75. And with traffic still light even at rush hour, it is possible that the owners may lower tolls even further in search of higher revenue.

After all, when the price was lowered by 45 percent last spring, it generated a 200 percent increase in volume three months later. If the same ratio applies again, lowering the toll another 25 percent would drive the daily volume up to 38,000 trips, and daily revenue up to nearly \$29,000.

The problem, of course, is that the same ratio usually does not apply at

every price point, which is why this pricing business is so tricky. . . .

Clifford Winston of the Brookings Institution and John Calfee of the American Enterprise Institute have considered the toll road's dilemma. . . .

Last year, the economists conducted an elaborate market test with 1,170 people across the country who were each presented with a series of options in which they were, in effect, asked to make a personal tradeoff between less commuting time and higher tolls.

In the end, they concluded that the people who placed the highest value on reducing their commuting time already had done so by finding public transportation, living closer to their work, or selecting jobs that allowed them to commute at off-peak hours.

Conversely, those who commuted significant distances had a higher tolerance for traffic congestion and were willing to pay only 20 percent of their hourly pay to save an hour of their time.

Overall, the Winston/Calfee findings help explain why the Greenway's original toll and volume projections were too high: By their reckoning, only commuters who earned at least \$30 an hour (about \$60,000 a year) would be willing to pay \$2 to save 20 minutes.

SOURCE: *The Washington Post*, October 24, 1996, p. E1.

$$\text{Income elasticity of demand} = \frac{\text{Percentage change in quantity demanded}}{\text{Percentage change in income}}.$$

As we discussed in Chapter 4, most goods are *normal goods*: Higher income raises quantity demanded. Because quantity demanded and income move in the same direction, normal goods have positive income elasticities. A few goods, such as bus

rides, are *inferior goods*: Higher income lowers the quantity demanded. Because quantity demanded and income move in opposite directions, inferior goods have negative income elasticities.

Even among normal goods, income elasticities vary substantially in size. Necessities, such as food and clothing, tend to have small income elasticities because consumers, regardless of how low their incomes, choose to buy some of these goods. Luxuries, such as caviar and furs, tend to have large income elasticities because consumers feel that they can do without these goods altogether if their income is too low.

cross-price elasticity of demand

a measure of how much the quantity demanded of one good responds to a change in the price of another good, computed as the percentage change in quantity demanded of the first good divided by the percentage change in the price of the second good

The Cross-Price Elasticity of Demand Economists use the **cross-price elasticity of demand** to measure how the quantity demanded of one good changes as the price of another good changes. It is calculated as the percentage change in quantity demanded of good 1 divided by the percentage change in the price of good 2. That is,

$$\text{Cross-price elasticity of demand} = \frac{\text{Percentage change in quantity demanded of good 1}}{\text{Percentage change in the price of good 2}}.$$

Whether the cross-price elasticity is a positive or negative number depends on whether the two goods are substitutes or complements. As we discussed in Chapter 4, substitutes are goods that are typically used in place of one another, such as hamburgers and hot dogs. An increase in hot dog prices induces people to grill hamburgers instead. Because the price of hot dogs and the quantity of hamburgers demanded move in the same direction, the cross-price elasticity is positive. Conversely, complements are goods that are typically used together, such as computers and software. In this case, the cross-price elasticity is negative, indicating that an increase in the price of computers reduces the quantity of software demanded.

QUICK QUIZ: Define the *price elasticity of demand*. ♦ Explain the relationship between total revenue and the price elasticity of demand.

THE ELASTICITY OF SUPPLY

When we discussed the determinants of supply in Chapter 4, we noted that sellers of a good increase the quantity supplied when the price of the good rises, when their input prices fall, or when their technology improves. To turn from qualitative to quantitative statements about supply, we once again use the concept of elasticity.

price elasticity of supply

a measure of how much the quantity supplied of a good responds to a change in the price of that good, computed as the percentage change in quantity supplied divided by the percentage change in price

THE PRICE ELASTICITY OF SUPPLY AND ITS DETERMINANTS

The law of supply states that higher prices raise the quantity supplied. The **price elasticity of supply** measures how much the quantity supplied responds to changes in the price. Supply of a good is said to be *elastic* if the quantity supplied

responds substantially to changes in the price. Supply is said to be *inelastic* if the quantity supplied responds only slightly to changes in the price.

The price elasticity of supply depends on the flexibility of sellers to change the amount of the good they produce. For example, beachfront land has an inelastic supply because it is almost impossible to produce more of it. By contrast, manufactured goods, such as books, cars, and televisions, have elastic supplies because the firms that produce them can run their factories longer in response to a higher price.

In most markets, a key determinant of the price elasticity of supply is the time period being considered. Supply is usually more elastic in the long run than in the short run. Over short periods of time, firms cannot easily change the size of their factories to make more or less of a good. Thus, in the short run, the quantity supplied is not very responsive to the price. By contrast, over longer periods, firms can build new factories or close old ones. In addition, new firms can enter a market, and old firms can shut down. Thus, in the long run, the quantity supplied can respond substantially to the price.

COMPUTING THE PRICE ELASTICITY OF SUPPLY

Now that we have some idea about what the price elasticity of supply is, let's be more precise. Economists compute the price elasticity of supply as the percentage change in the quantity supplied divided by the percentage change in the price. That is,

$$\text{Price elasticity of supply} = \frac{\text{Percentage change in quantity supplied}}{\text{Percentage change in price}}.$$

For example, suppose that an increase in the price of milk from \$2.85 to \$3.15 a gallon raises the amount that dairy farmers produce from 9,000 to 11,000 gallons per month. Using the midpoint method, we calculate the percentage change in price as

$$\text{Percentage change in price} = (3.15 - 2.85)/3.00 \times 100 = 10 \text{ percent}.$$

Similarly, we calculate the percentage change in quantity supplied as

$$\begin{aligned}\text{Percentage change in quantity supplied} &= (11,000 - 9,000)/10,000 \times 100 \\ &= 20 \text{ percent}.\end{aligned}$$

In this case, the price elasticity of supply is

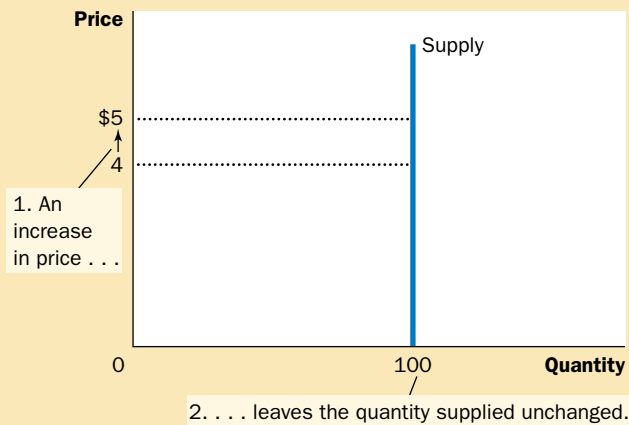
$$\text{Price elasticity of supply} = \frac{20 \text{ percent}}{10 \text{ percent}} = 2.0.$$

In this example, the elasticity of 2 reflects the fact that the quantity supplied moves proportionately twice as much as the price.

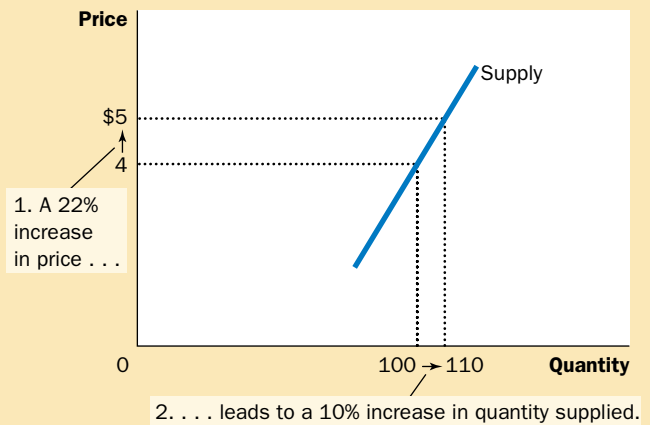
THE VARIETY OF SUPPLY CURVES

Because the price elasticity of supply measures the responsiveness of quantity supplied to the price, it is reflected in the appearance of the supply curve. Figure 5-6 shows five cases. In the extreme case of a zero elasticity, supply is *perfectly inelastic*,

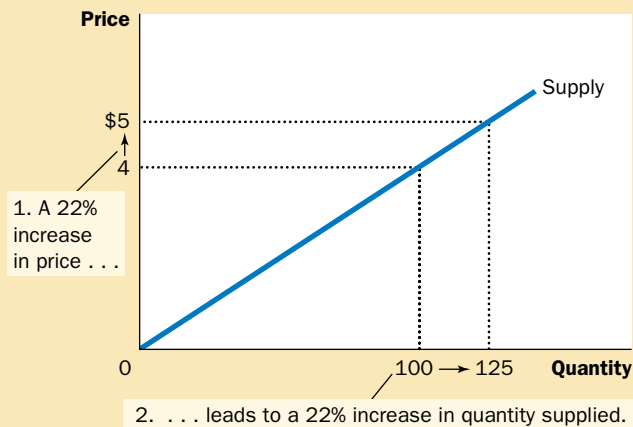
(a) Perfectly Inelastic Supply: Elasticity Equals 0



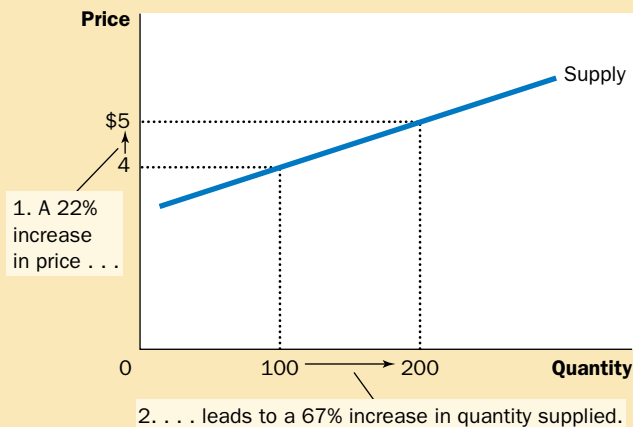
(b) Inelastic Supply: Elasticity Is Less Than 1



(c) Unit Elastic Supply: Elasticity Equals 1



(d) Elastic Supply: Elasticity Is Greater Than 1



(e) Perfectly Elastic Supply: Elasticity Equals Infinity

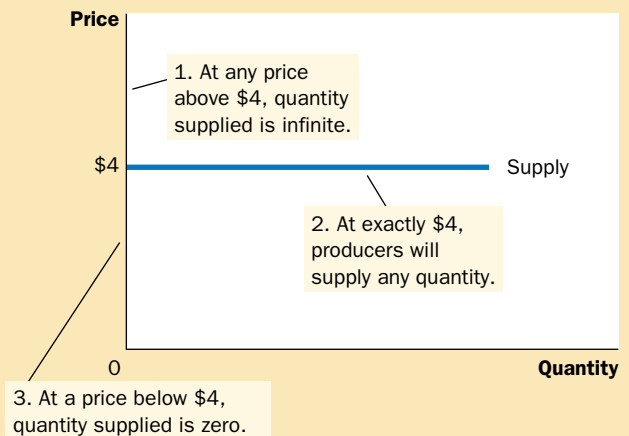
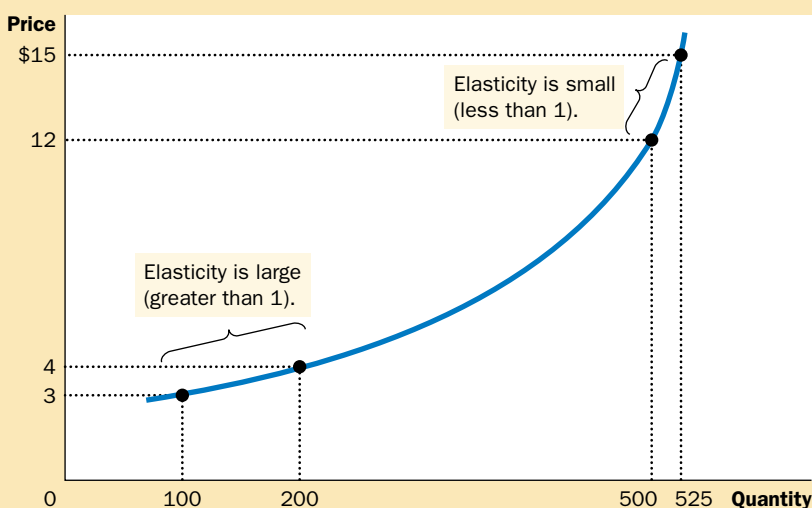


Figure 5-6

THE PRICE ELASTICITY OF SUPPLY. The price elasticity of supply determines whether the supply curve is steep or flat. Note that all percentage changes are calculated using the midpoint method.

**Figure 5-7**

HOW THE PRICE ELASTICITY OF SUPPLY CAN VARY. Because firms often have a maximum capacity for production, the elasticity of supply may be very high at low levels of quantity supplied and very low at high levels of quantity supplied. Here, an increase in price from \$3 to \$4 increases the quantity supplied from 100 to 200. Because the increase in quantity supplied of 67 percent is larger than the increase in price of 29 percent, the supply curve is elastic in this range. By contrast, when the price rises from \$12 to \$15, the quantity supplied rises only from 500 to 525. Because the increase in quantity supplied of 5 percent is smaller than the increase in price of 22 percent, the supply curve is inelastic in this range.

and the supply curve is vertical. In this case, the quantity supplied is the same regardless of the price. As the elasticity rises, the supply curve gets flatter, which shows that the quantity supplied responds more to changes in the price. At the opposite extreme, supply is *perfectly elastic*. This occurs as the price elasticity of supply approaches infinity and the supply curve becomes horizontal, meaning that very small changes in the price lead to very large changes in the quantity supplied.

In some markets, the elasticity of supply is not constant but varies over the supply curve. Figure 5-7 shows a typical case for an industry in which firms have factories with a limited capacity for production. For low levels of quantity supplied, the elasticity of supply is high, indicating that firms respond substantially to changes in the price. In this region, firms have capacity for production that is not being used, such as plants and equipment sitting idle for all or part of the day. Small increases in price make it profitable for firms to begin using this idle capacity. As the quantity supplied rises, firms begin to reach capacity. Once capacity is fully used, increasing production further requires the construction of new plants. To induce firms to incur this extra expense, the price must rise substantially, so supply becomes less elastic.

Figure 5-7 presents a numerical example of this phenomenon. When the price rises from \$3 to \$4 (a 29 percent increase, according to the midpoint method), the quantity supplied rises from 100 to 200 (a 67 percent increase). Because quantity supplied moves proportionately more than the price, the supply curve has elasticity greater than 1. By contrast, when the price rises from \$12 to \$15 (a 22 percent increase), the quantity supplied rises from 500 to 525 (a 5 percent increase). In this case, quantity supplied moves proportionately less than the price, so the elasticity is less than 1.

QUICK QUIZ: Define the *price elasticity of supply*. ♦ Explain why the the price elasticity of supply might be different in the long run than in the short run.

THREE APPLICATIONS OF SUPPLY, DEMAND, AND ELASTICITY

Can good news for farming be bad news for farmers? Why did the Organization of Petroleum Exporting Countries (OPEC) fail to keep the price of oil high? Does drug interdiction increase or decrease drug-related crime? At first, these questions might seem to have little in common. Yet all three questions are about markets, and all markets are subject to the forces of supply and demand. Here we apply the versatile tools of supply, demand, and elasticity to answer these seemingly complex questions.

CAN GOOD NEWS FOR FARMING BE BAD NEWS FOR FARMERS?

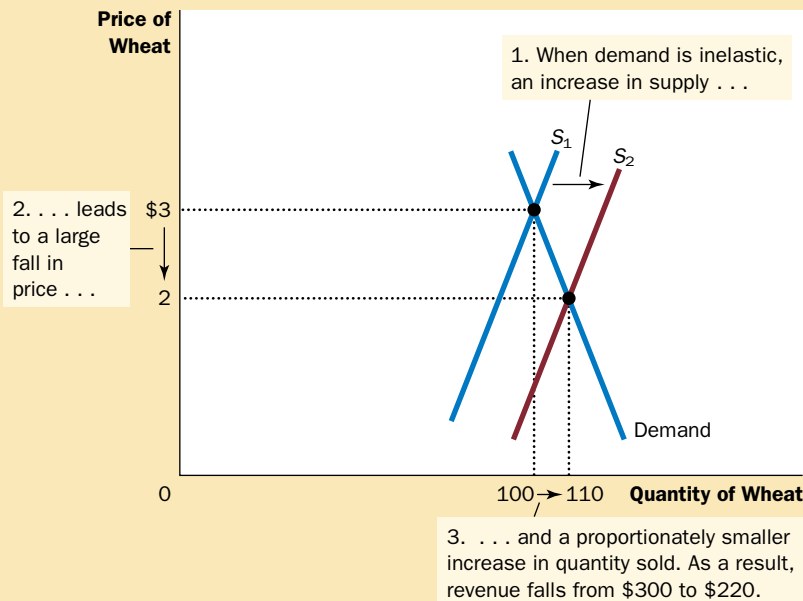
Let's now return to the question posed at the beginning of this chapter: What happens to wheat farmers and the market for wheat when university agronomists discover a new wheat hybrid that is more productive than existing varieties? Recall from Chapter 4 that we answer such questions in three steps. First, we examine whether the supply curve or demand curve shifts. Second, we consider which direction the curve shifts. Third, we use the supply-and-demand diagram to see how the market equilibrium changes.

In this case, the discovery of the new hybrid affects the supply curve. Because the hybrid increases the amount of wheat that can be produced on each acre of land, farmers are now willing to supply more wheat at any given price. In other words, the supply curve shifts to the right. The demand curve remains the same because consumers' desire to buy wheat products at any given price is not affected by the introduction of a new hybrid. Figure 5-8 shows an example of such a change. When the supply curve shifts from S_1 to S_2 , the quantity of wheat sold increases from 100 to 110, and the price of wheat falls from \$3 to \$2.

But does this discovery make farmers better off? As a first cut to answering this question, consider what happens to the total revenue received by farmers. Farmers' total revenue is $P \times Q$, the price of the wheat times the quantity sold. The discovery affects farmers in two conflicting ways. The hybrid allows farmers to produce more wheat (Q rises), but now each bushel of wheat sells for less (P falls).

Whether total revenue rises or falls depends on the elasticity of demand. In practice, the demand for basic foodstuffs such as wheat is usually inelastic, for these items are relatively inexpensive and have few good substitutes. When the demand curve is inelastic, as it is in Figure 5-8, a decrease in price causes total revenue to fall. You can see this in the figure: The price of wheat falls substantially, whereas the quantity of wheat sold rises only slightly. Total revenue falls from \$300 to \$220. Thus, the discovery of the new hybrid lowers the total revenue that farmers receive for the sale of their crops.

If farmers are made worse off by the discovery of this new hybrid, why do they adopt it? The answer to this question goes to the heart of how competitive markets work. Because each farmer is a small part of the market for wheat, he or she takes the price of wheat as given. For any given price of wheat, it is better to

Figure 5-8

AN INCREASE IN SUPPLY IN THE MARKET FOR WHEAT. When an advance in farm technology increases the supply of wheat from S_1 to S_2 , the price of wheat falls. Because the demand for wheat is inelastic, the increase in the quantity sold from 100 to 110 is proportionately smaller than the decrease in the price from \$3 to \$2. As a result, farmers' total revenue falls from \$300 ($\$3 \times 100$) to \$220 ($\$2 \times 110$).

use the new hybrid in order to produce and sell more wheat. Yet when all farmers do this, the supply of wheat rises, the price falls, and farmers are worse off.

Although this example may at first seem only hypothetical, in fact it helps to explain a major change in the U.S. economy over the past century. Two hundred years ago, most Americans lived on farms. Knowledge about farm methods was sufficiently primitive that most of us had to be farmers to produce enough food. Yet, over time, advances in farm technology increased the amount of food that each farmer could produce. This increase in food supply, together with inelastic food demand, caused farm revenues to fall, which in turn encouraged people to leave farming.

A few numbers show the magnitude of this historic change. As recently as 1950, there were 10 million people working on farms in the United States, representing 17 percent of the labor force. In 1998, fewer than 3 million people worked on farms, or 2 percent of the labor force. This change coincided with tremendous advances in farm productivity: Despite the 70 percent drop in the number of farmers, U.S. farms produced more than twice the output of crops and livestock in 1998 as they did in 1950.

This analysis of the market for farm products also helps to explain a seeming paradox of public policy: Certain farm programs try to help farmers by inducing them not to plant crops on all of their land. Why do these programs do this? Their purpose is to reduce the supply of farm products and thereby raise prices. With inelastic demand for their products, farmers as a group receive greater total revenue if they supply a smaller crop to the market. No single farmer would choose to leave his land fallow on his own because each takes the market price as given. But if all farmers do so together, each of them can be better off.

Doonesbury

by G.B. TRUDEAU



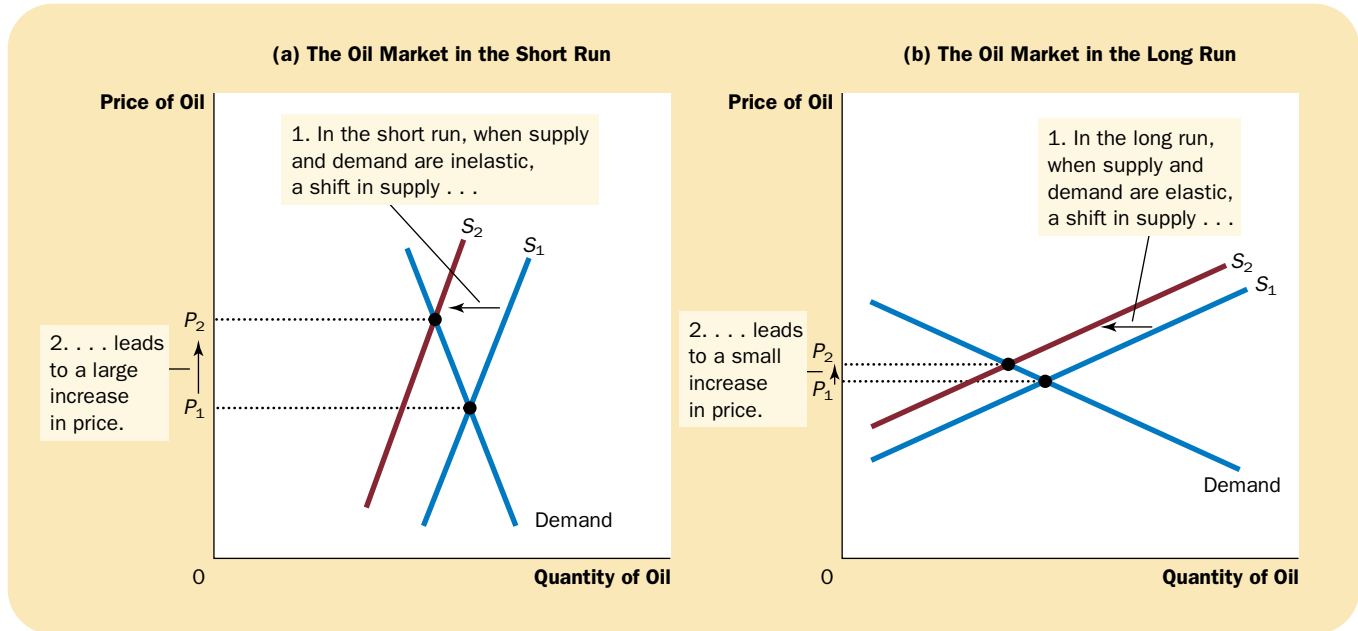
When analyzing the effects of farm technology or farm policy, it is important to keep in mind that what is good for farmers is not necessarily good for society as a whole. Improvement in farm technology can be bad for farmers who become increasingly unnecessary, but it is surely good for consumers who pay less for food. Similarly, a policy aimed at reducing the supply of farm products may raise the incomes of farmers, but it does so at the expense of consumers.

WHY DID OPEC FAIL TO KEEP THE PRICE OF OIL HIGH?

Many of the most disruptive events for the world's economies over the past several decades have originated in the world market for oil. In the 1970s members of the Organization of Petroleum Exporting Countries (OPEC) decided to raise the world price of oil in order to increase their incomes. These countries accomplished this goal by jointly reducing the amount of oil they supplied. From 1973 to 1974, the price of oil (adjusted for overall inflation) rose more than 50 percent. Then, a few years later, OPEC did the same thing again. The price of oil rose 14 percent in 1979, followed by 34 percent in 1980, and another 34 percent in 1981.

Yet OPEC found it difficult to maintain a high price. From 1982 to 1985, the price of oil steadily declined at about 10 percent per year. Dissatisfaction and disarray soon prevailed among the OPEC countries. In 1986 cooperation among OPEC members completely broke down, and the price of oil plunged 45 percent. In 1990 the price of oil (adjusted for overall inflation) was back to where it began in 1970, and it has stayed at that low level throughout most of the 1990s.

This episode shows how supply and demand can behave differently in the short run and in the long run. In the short run, both the supply and demand for oil are relatively inelastic. Supply is inelastic because the quantity of known oil reserves and the capacity for oil extraction cannot be changed quickly. Demand is inelastic because buying habits do not respond immediately to changes in price. Many drivers with old gas-guzzling cars, for instance, will just pay the higher



A REDUCTION IN SUPPLY IN THE WORLD MARKET FOR OIL. When the supply of oil falls, the response depends on the time horizon. In the short run, supply and demand are relatively inelastic, as in panel (a). Thus, when the supply curve shifts from S_1 to S_2 , the price rises substantially. By contrast, in the long run, supply and demand are relatively elastic, as in panel (b). In this case, the same size shift in the supply curve (S_1 to S_2) causes a smaller increase in the price.

Figure 5-9

price. Thus, as panel (a) of Figure 5-9 shows, the short-run supply and demand curves are steep. When the supply of oil shifts from S_1 to S_2 , the price increase from P_1 to P_2 is large.

The situation is very different in the long run. Over long periods of time, producers of oil outside of OPEC respond to high prices by increasing oil exploration and by building new extraction capacity. Consumers respond with greater conservation, for instance by replacing old inefficient cars with newer efficient ones. Thus, as panel (b) of Figure 5-9 shows, the long-run supply and demand curves are more elastic. In the long run, the shift in the supply curve from S_1 to S_2 causes a much smaller increase in the price.

This analysis shows why OPEC succeeded in maintaining a high price of oil only in the short run. When OPEC countries agreed to reduce their production of oil, they shifted the supply curve to the left. Even though each OPEC member sold less oil, the price rose by so much in the short run that OPEC incomes rose. By contrast, in the long run when supply and demand are more elastic, the same reduction in supply, measured by the horizontal shift in the supply curve, caused a smaller increase in the price. Thus, OPEC's coordinated reduction in supply proved less profitable in the long run.

OPEC still exists today, and it has from time to time succeeded at reducing supply and raising prices. But the price of oil (adjusted for overall inflation) has

never returned to the peak reached in 1981. The cartel now seems to understand that raising prices is easier in the short run than in the long run.

DOES DRUG INTERDICTION INCREASE OR DECREASE DRUG-RELATED CRIME?

A persistent problem facing our society is the use of illegal drugs, such as heroin, cocaine, and crack. Drug use has several adverse effects. One is that drug dependency can ruin the lives of drug users and their families. Another is that drug addicts often turn to robbery and other violent crimes to obtain the money needed to support their habit. To discourage the use of illegal drugs, the U.S. government devotes billions of dollars each year to reduce the flow of drugs into the country. Let's use the tools of supply and demand to examine this policy of drug interdiction.

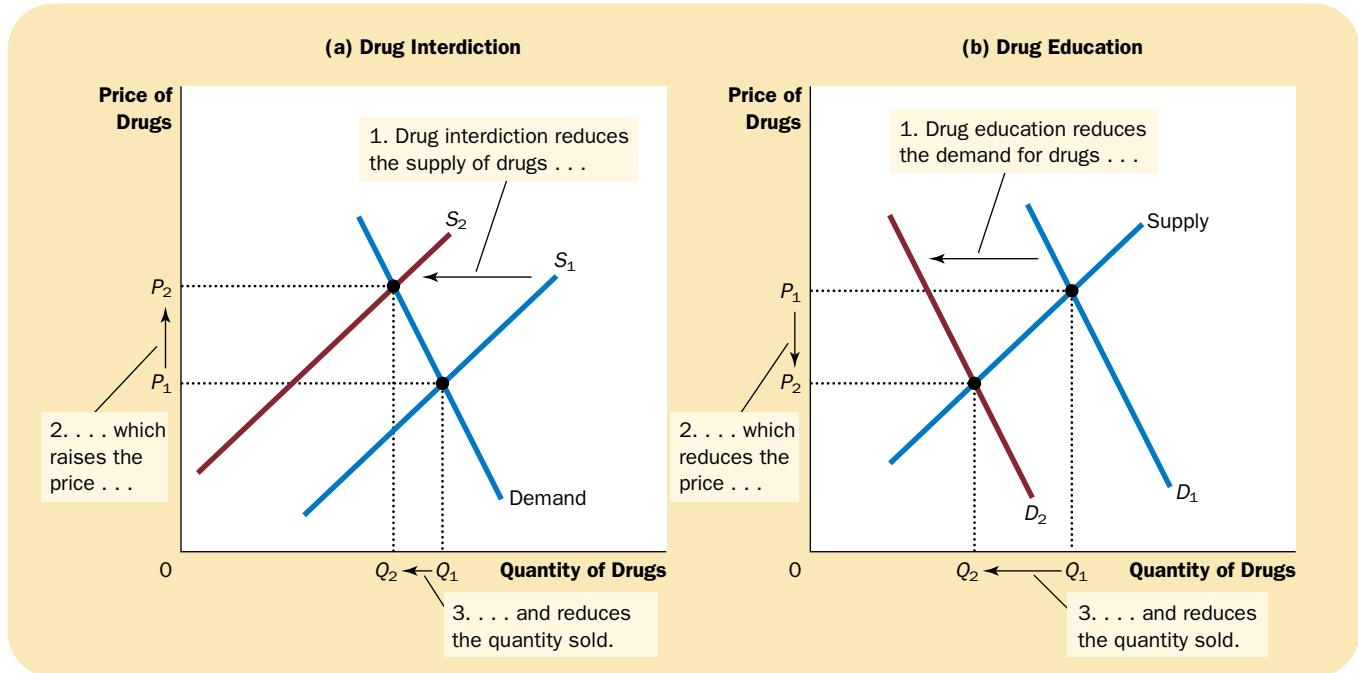
Suppose the government increases the number of federal agents devoted to the war on drugs. What happens in the market for illegal drugs? As is usual, we answer this question in three steps. First, we consider whether the supply curve or demand curve shifts. Second, we consider the direction of the shift. Third, we see how the shift affects the equilibrium price and quantity.

Although the purpose of drug interdiction is to reduce drug use, its direct impact is on the sellers of drugs rather than the buyers. When the government stops some drugs from entering the country and arrests more smugglers, it raises the cost of selling drugs and, therefore, reduces the quantity of drugs supplied at any given price. The demand for drugs—the amount buyers want at any given price—is not changed. As panel (a) of Figure 5-10 shows, interdiction shifts the supply curve to the left from S_1 to S_2 and leaves the demand curve the same. The equilibrium price of drugs rises from P_1 to P_2 , and the equilibrium quantity falls from Q_1 to Q_2 . The fall in the equilibrium quantity shows that drug interdiction does reduce drug use.

But what about the amount of drug-related crime? To answer this question, consider the total amount that drug users pay for the drugs they buy. Because few drug addicts are likely to break their destructive habits in response to a higher price, it is likely that the demand for drugs is inelastic, as it is drawn in the figure. If demand is inelastic, then an increase in price raises total revenue in the drug market. That is, because drug interdiction raises the price of drugs proportionately more than it reduces drug use, it raises the total amount of money that drug users pay for drugs. Addicts who already had to steal to support their habits would have an even greater need for quick cash. Thus, drug interdiction could increase drug-related crime.

Because of this adverse effect of drug interdiction, some analysts argue for alternative approaches to the drug problem. Rather than trying to reduce the supply of drugs, policymakers might try to reduce the demand by pursuing a policy of drug education. Successful drug education has the effects shown in panel (b) of Figure 5-10. The demand curve shifts to the left from D_1 to D_2 . As a result, the equilibrium quantity falls from Q_1 to Q_2 , and the equilibrium price falls from P_1 to P_2 . Total revenue, which is price times quantity, also falls. Thus, in contrast to drug interdiction, drug education can reduce both drug use and drug-related crime.

Advocates of drug interdiction might argue that the effects of this policy are different in the long run than in the short run, because the elasticity of demand may depend on the time horizon. The demand for drugs is probably inelastic over



POLICIES TO REDUCE THE USE OF ILLEGAL DRUGS. Drug interdiction reduces the supply of drugs from S_1 to S_2 , as in panel (a). If the demand for drugs is inelastic, then the total amount paid by drug users rises, even as the amount of drug use falls. By contrast, drug education reduces the demand for drugs from D_1 to D_2 , as in panel (b). Because both price and quantity fall, the amount paid by drug users falls.

Figure 5-10

short periods of time because higher prices do not substantially affect drug use by established addicts. But demand may be more elastic over longer periods of time because higher prices would discourage experimentation with drugs among the young and, over time, lead to fewer drug addicts. In this case, drug interdiction would increase drug-related crime in the short run while decreasing it in the long run.

QUICK QUIZ: How might a drought that destroys half of all farm crops be good for farmers? If such a drought is good for farmers, why don't farmers destroy their own crops in the absence of a drought?

CONCLUSION

According to an old quip, even a parrot can become an economist simply by learning to say "supply and demand." These last two chapters should have convinced you that there is much truth in this statement. The tools of supply and demand allow you to analyze many of the most important events and policies that shape

the economy. You are now well on your way to becoming an economist (or, at least, a well-educated parrot).

Summary

- ◆ The price elasticity of demand measures how much the quantity demanded responds to changes in the price. Demand tends to be more elastic if the good is a luxury rather than a necessity, if close substitutes are available, if the market is narrowly defined, or if buyers have substantial time to react to a price change.
- ◆ The price elasticity of demand is calculated as the percentage change in quantity demanded divided by the percentage change in price. If the elasticity is less than 1, so that quantity demanded moves proportionately less than the price, demand is said to be inelastic. If the elasticity is greater than 1, so that quantity demanded moves proportionately more than the price, demand is said to be elastic.
- ◆ Total revenue, the total amount paid for a good, equals the price of the good times the quantity sold. For inelastic demand curves, total revenue rises as price rises. For elastic demand curves, total revenue falls as price rises.
- ◆ The income elasticity of demand measures how much the quantity demanded responds to changes in consumers' income. The cross-price elasticity of demand measures how much the quantity demanded of one good responds to the price of another good.
- ◆ The price elasticity of supply measures how much the quantity supplied responds to changes in the price. This elasticity often depends on the time horizon under consideration. In most markets, supply is more elastic in the long run than in the short run.
- ◆ The price elasticity of supply is calculated as the percentage change in quantity supplied divided by the percentage change in price. If the elasticity is less than 1, so that quantity supplied moves proportionately less than the price, supply is said to be inelastic. If the elasticity is greater than 1, so that quantity supplied moves proportionately more than the price, supply is said to be elastic.
- ◆ The tools of supply and demand can be applied in many different kinds of markets. This chapter uses them to analyze the market for wheat, the market for oil, and the market for illegal drugs.

Key Concepts

elasticity, p. 94

price elasticity of demand, p. 94

total revenue, p. 98

income elasticity of demand, p. 102

cross-price elasticity of demand, p. 104

price elasticity of supply, p. 104

Questions for Review

1. Define the price elasticity of demand and the income elasticity of demand.
2. List and explain some of the determinants of the price elasticity of demand.
3. If the elasticity is greater than 1, is demand elastic or inelastic? If the elasticity equals 0, is demand perfectly elastic or perfectly inelastic?
4. On a supply-and-demand diagram, show equilibrium price, equilibrium quantity, and the total revenue received by producers.
5. If demand is elastic, how will an increase in price change total revenue? Explain.
6. What do we call a good whose income elasticity is less than 0?
7. How is the price elasticity of supply calculated? Explain what this measures.
8. What is the price elasticity of supply of Picasso paintings?
9. Is the price elasticity of supply usually larger in the short run or in the long run? Why?
10. In the 1970s, OPEC caused a dramatic increase in the price of oil. What prevented it from maintaining this high price through the 1980s?

Problems and Applications

- For each of the following pairs of goods, which good would you expect to have more elastic demand and why?
 - required textbooks or mystery novels
 - Beethoven recordings or classical music recordings in general
 - heating oil during the next six months or heating oil during the next five years
 - root beer or water
- Suppose that business travelers and vacationers have the following demand for airline tickets from New York to Boston:

PRICE	QUANTITY DEMANDED (BUSINESS TRAVELERS)	QUANTITY DEMANDED (VACATIONERS)
\$150	2,100	1,000
200	2,000	800
250	1,900	600
300	1,800	400

- As the price of tickets rises from \$200 to \$250, what is the price elasticity of demand for (i) business travelers and (ii) vacationers? (Use the midpoint method in your calculations.)
 - Why might vacationers have a different elasticity than business travelers?
- Suppose that your demand schedule for compact discs is as follows:

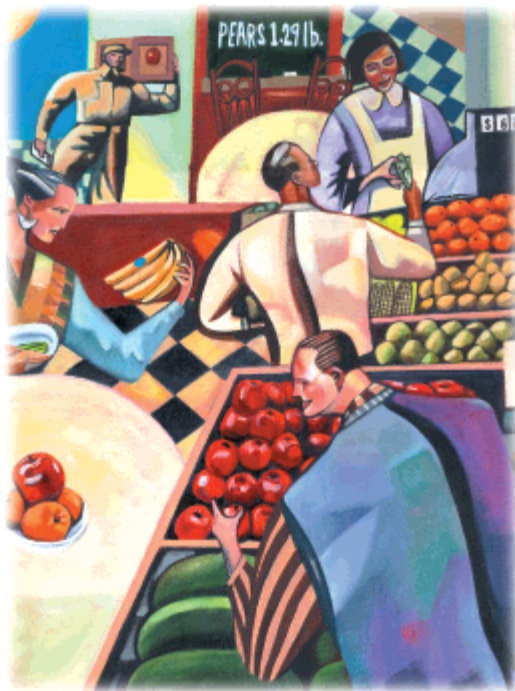
PRICE	QUANTITY DEMANDED (INCOME = \$10,000)	QUANTITY DEMANDED (INCOME = \$12,000)
\$ 8	40	50
10	32	45
12	24	30
14	16	20
16	8	12

- Use the midpoint method to calculate your price elasticity of demand as the price of compact discs increases from \$8 to \$10 if (i) your income is \$10,000, and (ii) your income is \$12,000.
 - Calculate your income elasticity of demand as your income increases from \$10,000 to \$12,000 if (i) the price is \$12, and (ii) the price is \$16.
- Emily has decided always to spend one-third of her income on clothing.
 - What is her income elasticity of clothing demand?
 - What is her price elasticity of clothing demand?
 - If Emily's tastes change and she decides to spend only one-fourth of her income on clothing, how does her demand curve change? What are her income elasticity and price elasticity now?
 - The New York Times* reported (Feb. 17, 1996, p. 25) that subway ridership declined after a fare increase: "There were nearly four million fewer riders in December 1995, the first full month after the price of a token increased 25 cents to \$1.50, than in the previous December, a 4.3 percent decline."
 - Use these data to estimate the price elasticity of demand for subway rides.
 - According to your estimate, what happens to the Transit Authority's revenue when the fare rises?
 - Why might your estimate of the elasticity be unreliable?
 - Two drivers—Tom and Jerry—each drive up to a gas station. Before looking at the price, each places an order. Tom says, "I'd like 10 gallons of gas." Jerry says, "I'd like \$10 worth of gas." What is each driver's price elasticity of demand?
 - Economists have observed that spending on restaurant meals declines more during economic downturns than does spending on food to be eaten at home. How might the concept of elasticity help to explain this phenomenon?
 - Consider public policy aimed at smoking.
 - Studies indicate that the price elasticity of demand for cigarettes is about 0.4. If a pack of cigarettes currently costs \$2 and the government wants to reduce smoking by 20 percent, by how much should it increase the price?
 - If the government permanently increases the price of cigarettes, will the policy have a larger effect on smoking one year from now or five years from now?
 - Studies also find that teenagers have a higher price elasticity than do adults. Why might this be true?
 - Would you expect the price elasticity of *demand* to be larger in the market for all ice cream or the market for vanilla ice cream? Would you expect the price elasticity of *supply* to be larger in the market for all ice cream or the market for vanilla ice cream? Be sure to explain your answers.
 - Pharmaceutical drugs have an inelastic demand, and computers have an elastic demand. Suppose that

technological advance doubles the supply of both products (that is, the quantity supplied at each price is twice what it was).

- a. What happens to the equilibrium price and quantity in each market?
 - b. Which product experiences a larger change in price?
 - c. Which product experiences a larger change in quantity?
 - d. What happens to total consumer spending on each product?
11. Beachfront resorts have an inelastic supply, and automobiles have an elastic supply. Suppose that a rise in population doubles the demand for both products (that is, the quantity demanded at each price is twice what it was).
 - a. What happens to the equilibrium price and quantity in each market?
 - b. Which product experiences a larger change in price?
 - c. Which product experiences a larger change in quantity?
 - d. What happens to total consumer spending on each product?
 12. Several years ago, flooding along the Missouri and Mississippi rivers destroyed thousands of acres of wheat.
 - a. Farmers whose crops were destroyed by the floods were much worse off, but farmers whose crops were not destroyed benefited from the floods. Why?
 - b. What information would you need about the market for wheat in order to assess whether farmers as a group were hurt or helped by the floods?
 13. Explain why the following might be true: A drought around the world raises the total revenue that farmers receive from the sale of grain, but a drought only in Kansas reduces the total revenue that Kansas farmers receive.
 14. Because better weather makes farmland more productive, farmland in regions with good weather conditions is more expensive than farmland in regions with bad weather conditions. Over time, however, as advances in technology have made all farmland more productive, the price of farmland (adjusted for overall inflation) has fallen. Use the concept of elasticity to explain why productivity and farmland prices are positively related across space but negatively related over time.

6



SUPPLY, DEMAND, AND GOVERNMENT POLICIES

Economists have two roles. As scientists, they develop and test theories to explain the world around them. As policy advisers, they use their theories to help change the world for the better. The focus of the preceding two chapters has been scientific. We have seen how supply and demand determine the price of a good and the quantity of the good sold. We have also seen how various events shift supply and demand and thereby change the equilibrium price and quantity.

This chapter offers our first look at policy. Here we analyze various types of government policy using only the tools of supply and demand. As you will see, the analysis yields some surprising insights. Policies often have effects that their architects did not intend or anticipate.

We begin by considering policies that directly control prices. For example, rent-control laws dictate a maximum rent that landlords may charge tenants. Minimum-wage laws dictate the lowest wage that firms may pay workers. Price controls are

IN THIS CHAPTER YOU WILL . . .

***Examine the effects
of government
policies that place
a ceiling on prices***

***Examine the effects
of government
policies that put a
floor under prices***

***Consider how a tax
on a good affects
the price of the
good and the
quantity sold***

***Learn that taxes
levied on buyers
and taxes levied on
sellers are
equivalent***

***See how the burden
of a tax is split
between buyers
and sellers***

usually enacted when policymakers believe that the market price of a good or service is unfair to buyers or sellers. Yet, as we will see, these policies can generate inequities of their own.

After our discussion of price controls, we next consider the impact of taxes. Policymakers use taxes both to influence market outcomes and to raise revenue for public purposes. Although the prevalence of taxes in our economy is obvious, their effects are not. For example, when the government levies a tax on the amount that firms pay their workers, do the firms or the workers bear the burden of the tax? The answer is not at all clear—until we apply the powerful tools of supply and demand.

CONTROLS ON PRICES

To see how price controls affect market outcomes, let's look once again at the market for ice cream. As we saw in Chapter 4, if ice cream is sold in a competitive market free of government regulation, the price of ice cream adjusts to balance supply and demand: At the equilibrium price, the quantity of ice cream that buyers want to buy exactly equals the quantity that sellers want to sell. To be concrete, suppose the equilibrium price is \$3 per cone.

Not everyone may be happy with the outcome of this free-market process. Let's say the American Association of Ice Cream Eaters complains that the \$3 price is too high for everyone to enjoy a cone a day (their recommended diet). Meanwhile, the National Organization of Ice Cream Makers complains that the \$3 price—the result of “cutthroat competition”—is depressing the incomes of its members. Each of these groups lobbies the government to pass laws that alter the market outcome by directly controlling prices.

Of course, because buyers of any good always want a lower price while sellers want a higher price, the interests of the two groups conflict. If the Ice Cream Eaters are successful in their lobbying, the government imposes a legal maximum on the price at which ice cream can be sold. Because the price is not allowed to rise above this level, the legislated maximum is called a **price ceiling**. By contrast, if the Ice Cream Makers are successful, the government imposes a legal minimum on the price. Because the price cannot fall below this level, the legislated minimum is called a **price floor**. Let us consider the effects of these policies in turn.

price ceiling

a legal maximum on the price at which a good can be sold

price floor

a legal minimum on the price at which a good can be sold

HOW PRICE CEILINGS AFFECT MARKET OUTCOMES

When the government, moved by the complaints of the Ice Cream Eaters, imposes a price ceiling on the market for ice cream, two outcomes are possible. In panel (a) of Figure 6-1, the government imposes a price ceiling of \$4 per cone. In this case, because the price that balances supply and demand (\$3) is below the ceiling, the price ceiling is *not binding*. Market forces naturally move the economy to the equilibrium, and the price ceiling has no effect.

Panel (b) of Figure 6-1 shows the other, more interesting, possibility. In this case, the government imposes a price ceiling of \$2 per cone. Because the equilibrium price of \$3 is above the price ceiling, the ceiling is a *binding constraint* on the market.

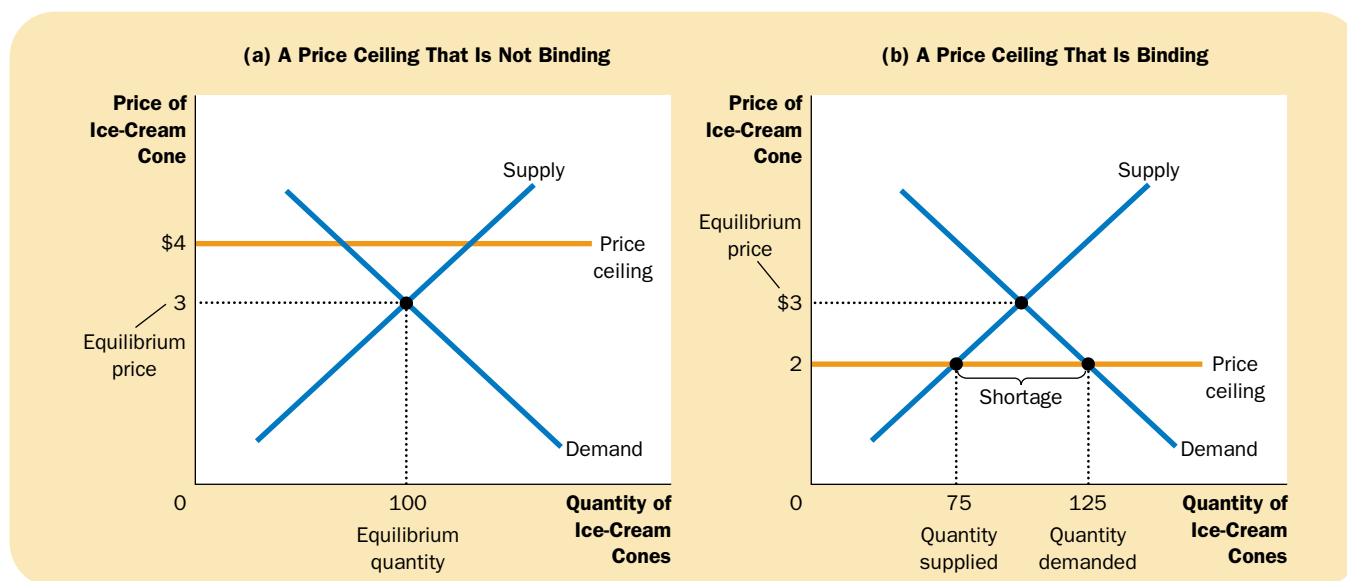


Figure 6-1

A MARKET WITH A PRICE CEILING. In panel (a), the government imposes a price ceiling of \$4. Because the price ceiling is above the equilibrium price of \$3, the price ceiling has no effect, and the market can reach the equilibrium of supply and demand. In this equilibrium, quantity supplied and quantity demanded both equal 100 cones. In panel (b), the government imposes a price ceiling of \$2. Because the price ceiling is below the equilibrium price of \$3, the market price equals \$2. At this price, 125 cones are demanded and only 75 are supplied, so there is a shortage of 50 cones.

The forces of supply and demand tend to move the price toward the equilibrium price, but when the market price hits the ceiling, it can rise no further. Thus, the market price equals the price ceiling. At this price, the quantity of ice cream demanded (125 cones in the figure) exceeds the quantity supplied (75 cones). There is a shortage of ice cream, so some people who want to buy ice cream at the going price are unable to.

When a shortage of ice cream develops because of this price ceiling, some mechanism for rationing ice cream will naturally develop. The mechanism could be long lines: Buyers who are willing to arrive early and wait in line get a cone, while those unwilling to wait do not. Alternatively, sellers could ration ice cream according to their own personal biases, selling it only to friends, relatives, or members of their own racial or ethnic group. Notice that even though the price ceiling was motivated by a desire to help buyers of ice cream, not all buyers benefit from the policy. Some buyers do get to pay a lower price, although they may have to wait in line to do so, but other buyers cannot get any ice cream at all.

This example in the market for ice cream shows a general result: *When the government imposes a binding price ceiling on a competitive market, a shortage of the good arises, and sellers must ration the scarce goods among the large number of potential buyers.* The rationing mechanisms that develop under price ceilings are rarely desirable. Long lines are inefficient, because they waste buyers' time. Discrimination according to seller bias is both inefficient (because the good does not go to the buyer who values it most highly) and potentially unfair. By contrast, the rationing mechanism



WHO IS RESPONSIBLE FOR THIS—OPEC OR U.S. LAWMAKERS?

in a free, competitive market is both efficient and impersonal. When the market for ice cream reaches its equilibrium, anyone who wants to pay the market price can get a cone. Free markets ration goods with prices.

CASE STUDY LINES AT THE GAS PUMP

As we discussed in the preceding chapter, in 1973 the Organization of Petroleum Exporting Countries (OPEC) raised the price of crude oil in world oil markets. Because crude oil is the major input used to make gasoline, the higher oil prices reduced the supply of gasoline. Long lines at gas stations became commonplace, and motorists often had to wait for hours to buy only a few gallons of gas.

What was responsible for the long gas lines? Most people blame OPEC. Surely, if OPEC had not raised the price of crude oil, the shortage of gasoline would not have occurred. Yet economists blame government regulations that limited the price oil companies could charge for gasoline.

Figure 6-2 shows what happened. As shown in panel (a), before OPEC raised the price of crude oil, the equilibrium price of gasoline P_1 was below the price ceiling. The price regulation, therefore, had no effect. When the price of crude oil rose, however, the situation changed. The increase in the price of crude

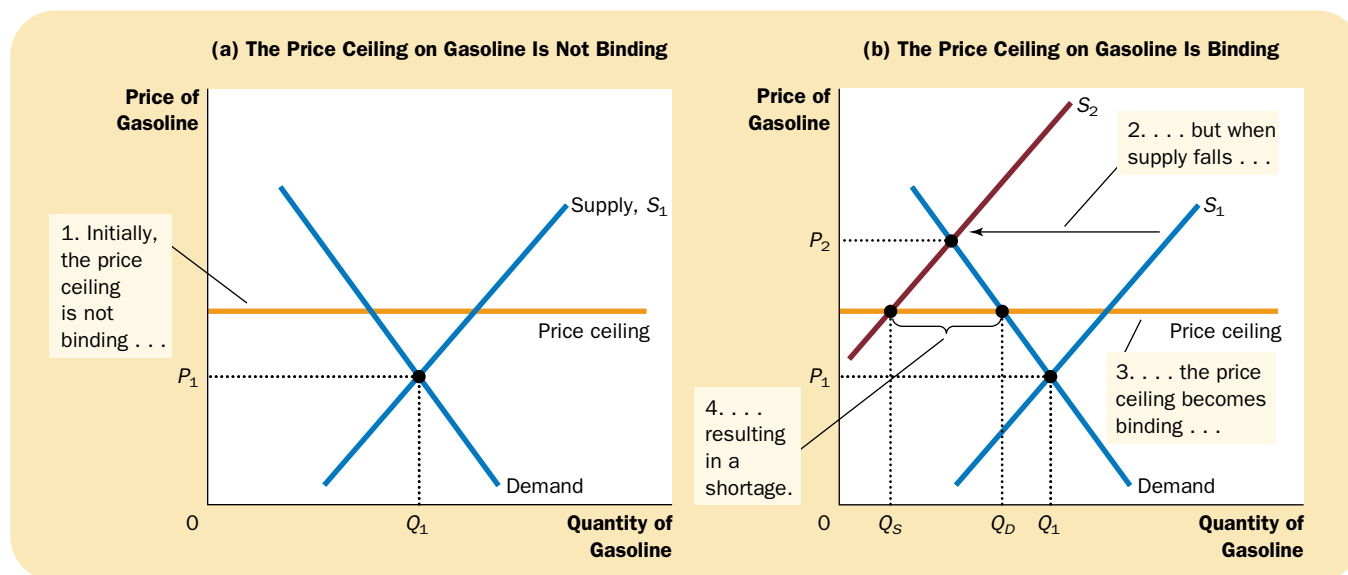


Figure 6-2

THE MARKET FOR GASOLINE WITH A PRICE CEILING. Panel (a) shows the gasoline market when the price ceiling is not binding because the equilibrium price, P_1 , is below the ceiling. Panel (b) shows the gasoline market after an increase in the price of crude oil (an input into making gasoline) shifts the supply curve to the left from S_1 to S_2 . In an unregulated market, the price would have risen from P_1 to P_2 . The price ceiling, however, prevents this from happening. At the binding price ceiling, consumers are willing to buy Q_D , but producers of gasoline are willing to sell only Q_S . The difference between quantity demanded and quantity supplied, $Q_D - Q_S$, measures the gasoline shortage.

IN THE NEWS*Does a Drought Need to Cause a Water Shortage?*

DURING THE SUMMER OF 1999, THE EAST coast of the United States experienced unusually little rain and a shortage of water. The following article suggests a way that the shortage could have been averted.

Trickle-Down Economics

BY TERRY L. ANDERSON AND
CLAY J. LANDRY

Water shortages are being blamed on the drought in the East, but that's giving Mother Nature a bum rap. Certainly the drought is the immediate cause, but the real culprit is regulations that don't allow markets and prices to equalize demand and supply.

The similarity between water and gasoline is instructive. The energy crisis of the 1970s, too, was blamed on nature's niggardly supply of oil, but in fact it was the actions of the Organization of Petroleum Exporting Countries, combined with price controls, that was the main cause of the shortages. . . .

Once again, regulators are responding to shortages—in this case of water—with controls and regulations rather than allowing the market to work. Cities are restricting water usage; some have even gone so far as to prohibit restaurants from serving water except if the customer asks for a glass. But although cities initially saw declines in water use, some are starting to report increases in consumption. This has prompted some police departments to collect lists of residents suspected of wasting water.

There's a better answer than sending out the cops. Market forces could ensure plentiful water availability even in drought years. Contrary to popular belief, the supply of water is no more fixed than the supply of oil. Like all resources, water supplies change in response to economic growth and to the price. In developing countries, despite population growth, the percentage of people with access to safe drinking water has increased to 74 percent in 1994 from 44 percent in 1980. Rising incomes have given those countries the wherewithal to supply potable water.

Supplies also increase when current users have an incentive to conserve their surplus in the marketplace. California's drought-emergency water bank illustrates this. The bank allows farmers to lease water from other users during dry spells. In 1991, the first year the bank was tried, when the price was \$125 per acre-foot (326,000 gallons), supply exceeded demand by two to one. That is,

many more people wanted to sell their water than wanted to buy.

Data from every corner of the world show that when cities raise the price of water by 10 percent, water use goes down by as much as 12 percent. When the price of agricultural water goes up 10 percent, usage goes down by 20 percent. . . .

Unfortunately, Eastern water users do not pay realistic prices for water. According to the American Water Works Association, only 2 percent of municipal water suppliers adjust prices seasonally. . . .

Even more egregious, Eastern water laws bar people from buying and selling water. Just as tradable pollution permits established under the Clean Air Act have encouraged polluters to find efficient ways to reduce emissions, tradable water rights can encourage conservation and increase supplies. It is mainly a matter of following the lead of Western water courts that have quantified water rights and Western legislatures that have allowed trades.

By making water a commodity and unleashing market forces, policymakers can ensure plentiful water supplies for all. New policies won't make droughts disappear, but they will ease the pain they impose by priming the invisible pump of water markets.

SOURCE: *The Wall Street Journal*, August 23, 1999, p. A14.

oil raised the cost of producing gasoline, and this reduced the supply of gasoline. As panel (b) shows, the supply curve shifted to the left from S_1 to S_2 . In an unregulated market, this shift in supply would have raised the equilibrium price of gasoline from P_1 to P_2 , and no shortage would have resulted. Instead, the price ceiling prevented the price from rising to the equilibrium level. At the

price ceiling, producers were willing to sell Q_S , and consumers were willing to buy Q_D . Thus, the shift in supply caused a severe shortage at the regulated price.

Eventually, the laws regulating the price of gasoline were repealed. Lawmakers came to understand that they were partly responsible for the many hours Americans lost waiting in line to buy gasoline. Today, when the price of crude oil changes, the price of gasoline can adjust to bring supply and demand into equilibrium.

CASE STUDY RENT CONTROL IN THE SHORT RUN AND LONG RUN

One common example of a price ceiling is rent control. In some cities, the local government places a ceiling on rents that landlords may charge their tenants. The goal of this policy is to help the poor by making housing more affordable. Economists often criticize rent control, arguing that it is a highly inefficient way to help the poor raise their standard of living. One economist called rent control “the best way to destroy a city, other than bombing.”

The adverse effects of rent control are less apparent to the general population because these effects occur over many years. In the short run, landlords have a fixed number of apartments to rent, and they cannot adjust this number quickly as market conditions change. Moreover, the number of people searching

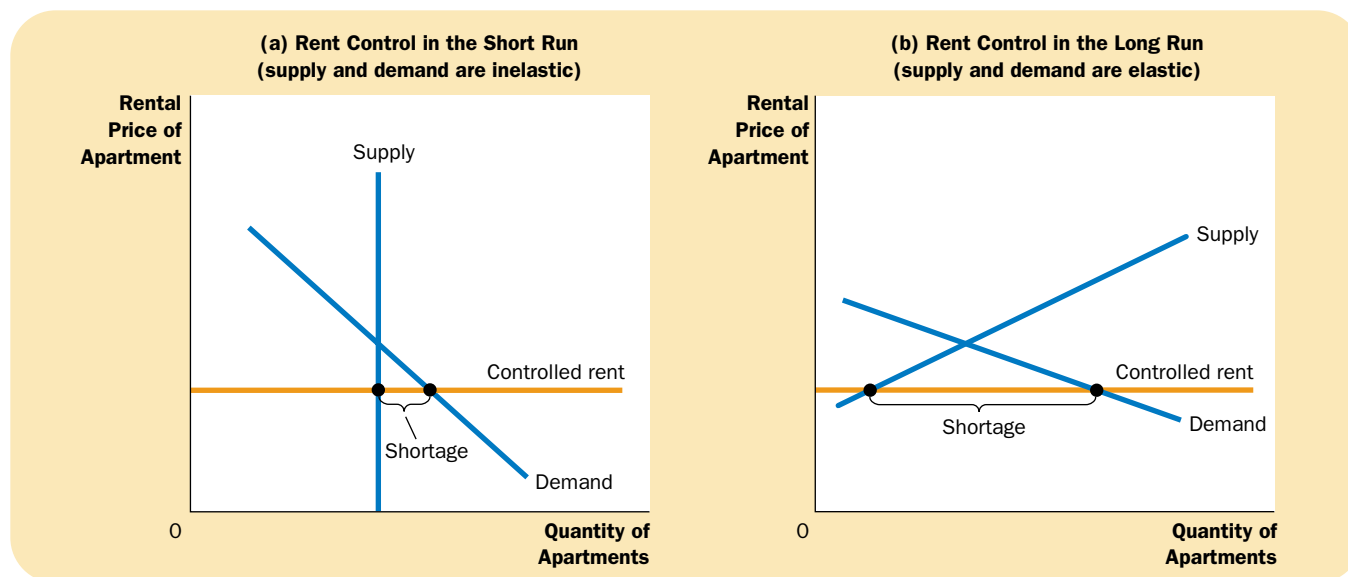


Figure 6-3

RENT CONTROL IN THE SHORT RUN AND IN THE LONG RUN. Panel (a) shows the short-run effects of rent control: Because the supply and demand for apartments are relatively inelastic, the price ceiling imposed by a rent-control law causes only a small shortage of housing. Panel (b) shows the long-run effects of rent control: Because the supply and demand for apartments are more elastic, rent control causes a large shortage.

for housing in a city may not be highly responsive to rents in the short run because people take time to adjust their housing arrangements. Therefore, the short-run supply and demand for housing are relatively inelastic.

Panel (a) of Figure 6-3 shows the short-run effects of rent control on the housing market. As with any price ceiling, rent control causes a shortage. Yet because supply and demand are inelastic in the short run, the initial shortage caused by rent control is small. The primary effect in the short run is to reduce rents.

The long-run story is very different because the buyers and sellers of rental housing respond more to market conditions as time passes. On the supply side, landlords respond to low rents by not building new apartments and by failing to maintain existing ones. On the demand side, low rents encourage people to find their own apartments (rather than living with their parents or sharing apartments with roommates) and induce more people to move into a city. Therefore, both supply and demand are more elastic in the long run.

Panel (b) of Figure 6-3 illustrates the housing market in the long run. When rent control depresses rents below the equilibrium level, the quantity of apartments supplied falls substantially, and the quantity of apartments demanded rises substantially. The result is a large shortage of housing.

In cities with rent control, landlords use various mechanisms to ration housing. Some landlords keep long waiting lists. Others give a preference to tenants without children. Still others discriminate on the basis of race. Sometimes, apartments are allocated to those willing to offer under-the-table payments to building superintendents. In essence, these bribes bring the total price of an apartment (including the bribe) closer to the equilibrium price.

To understand fully the effects of rent control, we have to remember one of the *Ten Principles of Economics* from Chapter 1: People respond to incentives. In free markets, landlords try to keep their buildings clean and safe because desirable apartments command higher prices. By contrast, when rent control creates shortages and waiting lists, landlords lose their incentive to be responsive to tenants' concerns. Why should a landlord spend his money to maintain and improve his property when people are waiting to get in as it is? In the end, tenants get lower rents, but they also get lower-quality housing.

Policymakers often react to the effects of rent control by imposing additional regulations. For example, there are laws that make racial discrimination in housing illegal and require landlords to provide minimally adequate living conditions. These laws, however, are difficult and costly to enforce. By contrast, when rent control is eliminated and a market for housing is regulated by the forces of competition, such laws are less necessary. In a free market, the price of housing adjusts to eliminate the shortages that give rise to undesirable landlord behavior.



HOW PRICE FLOORS AFFECT MARKET OUTCOMES

To examine the effects of another kind of government price control, let's return to the market for ice cream. Imagine now that the government is persuaded by the pleas of the National Organization of Ice Cream Makers. In this case, the government might institute a price floor. Price floors, like price ceilings, are an attempt by the government to maintain prices at other than equilibrium levels. Whereas a price ceiling places a legal maximum on prices, a price floor places a legal minimum.

IN THE NEWS*Rent Control in
New York City*

RENT CONTROL REMAINS A TOPIC OF HEATED debate in New York City, as the following article describes.

Threat to End Rent Control Stirs Up NYC

BY FRED KAPLAN

NEW YORK—One recent lunch hour at Shopsyin's, a neighborhood diner in Manhattan's West Village, conversation turned to the topic of the state Senate majority leader, Joseph L. Bruno. "If he ever shows his face around here, we'll string him up," a customer exclaimed. "The guy deserves death," another said matter-of-factly.

Rarely has so much venom been aimed at a figure so obscure as an Albany legislator, but all over New York City, thousands of otherwise fairly civilized citizens are throwing similar fits. For Bruno is threatening to take away their one holy fringe benefit—the eternal right to a rent-controlled apartment.

Massachusetts and California have abolished or scaled back their rent-control laws in recent years, but New York remains the last holdout, and on a scale that dwarfs that of the other cities.

About 2 million residents—more than a quarter of New York City's popu-

lation—live in apartments covered by regulations that severely limit how much a landlord can raise the rent and under what conditions a tenant or even a tenant's relatives can be evicted.

Tales are legion of wealthy movie stars, doctors, and stock brokers paying a pittance for palatial dwellings in the more fashionable neighborhoods of Manhattan.

Some of these tales were knocked off the books in 1993, when the state Legislature passed what many called "the Mia Farrow law"—in reference to the actress who was paying one-fifth the market price for a 10-room apartment on Central Park West. Still, the bill did not affect too many people. It lifted rent controls only from apartments going for more than \$2,000 a month, and only if the tenants's annual household income exceeded \$250,000 two years in a row.

Far more plentiful are the unaffected cases. An investment banker, who earns more than \$400,000 a year, pays \$1,500 a month for a three-bedroom apartment near Lincoln Center. A securities trader, making well over \$100,000 a year, pays \$800 a month for a one-bedroom on the Upper West Side. In both cases, the units would fetch at least three times as much if placed on the open market. . . .

But rent control helps more than the rich. A study by the city concludes that the average tenant of a rent-controlled apartment in New York City earns only \$20,000 a year. Tenants' groups say that ending controls would primarily raise the rents of those who can least afford to pay, resulting in wholesale eviction.

However, Paul Grogan, president of the Local Initiatives Support Corp., a private organization that finances low-

income housing, said, "In many poor neighborhoods, the landlord can't even get rents as high as the regulations allow." . . .

Few economists and policy analysts, even liberal ones, support rent control—not so much because it lets rich people pay far less than they can afford, but because it distorts the marketplace for everyone.

Frank Roconi, director of the Citizens Housing and Planning Council, a public-policy research organization that supports some government intervention in the real-estate market, spelled out "the classic case" of this distortion:

"There is an elderly couple, their kids are gone, they have a three-bedroom apartment, and they are paying \$400 a month. Down the hall, there is a young family with two kids living in a one-bedroom for \$1,000 a month. In a rational price system, the elderly couple would have an incentive to move to a smaller, cheaper apartment, leaving vacant a larger space for the young family."

Under the current system, though, if the elderly couple moves away, their children can claim the apartment at the same rent. Or, if it is left vacant, the landlord, by law, can charge only a few percentage points more than if the tenant had stayed.

Therefore, Roconi noted, "the landlord isn't going to let just anybody in. He's going to let his brother-in-law have the apartment or his accountant or someone willing to give him a bribe. There's a tremendous incentive for that apartment never to hit the open market."

SOURCE: *The Boston Globe*, April 28, 1997, p. A1.

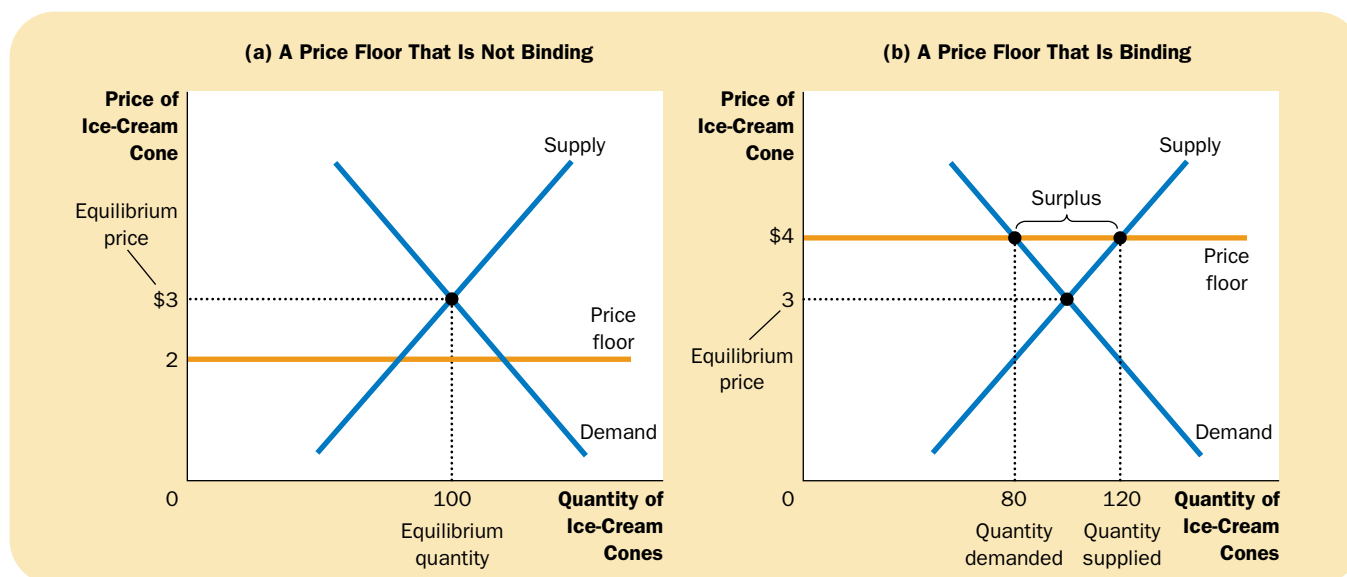


Figure 6-4

A MARKET WITH A PRICE FLOOR. In panel (a), the government imposes a price floor of \$2. Because this is below the equilibrium price of \$3, the price floor has no effect. The market price adjusts to balance supply and demand. At the equilibrium, quantity supplied and quantity demanded both equal 100 cones. In panel (b), the government imposes a price floor of \$4, which is above the equilibrium price of \$3. Therefore, the market price equals \$4. Because 120 cones are supplied at this price and only 80 are demanded, there is a surplus of 40 cones.

When the government imposes a price floor on the ice-cream market, two outcomes are possible. If the government imposes a price floor of \$2 per cone when the equilibrium price is \$3, we obtain the outcome in panel (a) of Figure 6-4. In this case, because the equilibrium price is above the floor, the price floor is not binding. Market forces naturally move the economy to the equilibrium, and the price floor has no effect.

Panel (b) of Figure 6-4 shows what happens when the government imposes a price floor of \$4 per cone. In this case, because the equilibrium price of \$3 is below the floor, the price floor is a binding constraint on the market. The forces of supply and demand tend to move the price toward the equilibrium price, but when the market price hits the floor, it can fall no further. The market price equals the price floor. At this floor, the quantity of ice cream supplied (120 cones) exceeds the quantity demanded (80 cones). Some people who want to sell ice cream at the going price are unable to. *Thus, a binding price floor causes a surplus.*

Just as price ceilings and shortages can lead to undesirable rationing mechanisms, so can price floors and surpluses. In the case of a price floor, some sellers are unable to sell all they want at the market price. The sellers who appeal to the personal biases of the buyers, perhaps due to racial or familial ties, are better able to sell their goods than those who do not. By contrast, in a free market, the price serves as the rationing mechanism, and sellers can sell all they want at the equilibrium price.

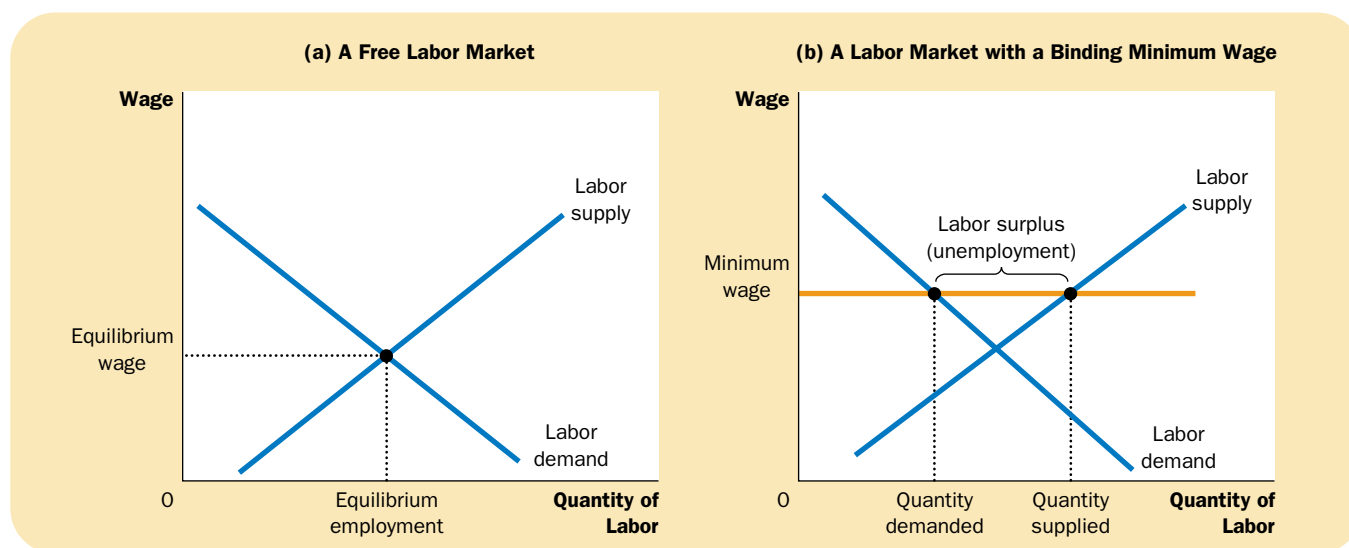
CASE STUDY THE MINIMUM WAGE

An important example of a price floor is the minimum wage. Minimum-wage laws dictate the lowest price for labor that any employer may pay. The U.S. Congress first instituted a minimum wage with the Fair Labor Standards Act of 1938 to ensure workers a minimally adequate standard of living. In 1999 the minimum wage according to federal law was \$5.15 per hour, and some state laws imposed higher minimum wages.

To examine the effects of a minimum wage, we must consider the market for labor. Panel (a) of Figure 6-5 shows the labor market which, like all markets, is subject to the forces of supply and demand. Workers determine the supply of labor, and firms determine the demand. If the government doesn't intervene, the wage normally adjusts to balance labor supply and labor demand.

Panel (b) of Figure 6-5 shows the labor market with a minimum wage. If the minimum wage is above the equilibrium level, as it is here, the quantity of labor supplied exceeds the quantity demanded. The result is unemployment. Thus, the minimum wage raises the incomes of those workers who have jobs, but it lowers the incomes of those workers who cannot find jobs.

To fully understand the minimum wage, keep in mind that the economy contains not a single labor market, but many labor markets for different types of workers. The impact of the minimum wage depends on the skill and experience of the worker. Workers with high skills and much experience are not affected, because their equilibrium wages are well above the minimum. For these workers, the minimum wage is not binding.

**Figure 6-5**

HOW THE MINIMUM WAGE AFFECTS THE LABOR MARKET. Panel (a) shows a labor market in which the wage adjusts to balance labor supply and labor demand. Panel (b) shows the impact of a binding minimum wage. Because the minimum wage is a price floor, it causes a surplus: The quantity of labor supplied exceeds the quantity demanded. The result is unemployment.

The minimum wage has its greatest impact on the market for teenage labor. The equilibrium wages of teenagers are low because teenagers are among the least skilled and least experienced members of the labor force. In addition, teenagers are often willing to accept a lower wage in exchange for on-the-job training. (Some teenagers are willing to work as “interns” for no pay at all. Because internships pay nothing, however, the minimum wage does not apply to them. If it did, these jobs might not exist.) As a result, the minimum wage is more often binding for teenagers than for other members of the labor force.

Many economists have studied how minimum-wage laws affect the teenage labor market. These researchers compare the changes in the minimum wage over time with the changes in teenage employment. Although there is some debate about how much the minimum wage affects employment, the typical study finds that a 10 percent increase in the minimum wage depresses teenage employment between 1 and 3 percent. In interpreting this estimate, note that a 10 percent increase in the minimum wage does not raise the average wage of teenagers by 10 percent. A change in the law does not directly affect those teenagers who are already paid well above the minimum, and enforcement of minimum-wage laws is not perfect. Thus, the estimated drop in employment of 1 to 3 percent is significant.

In addition to altering the quantity of labor demanded, the minimum wage also alters the quantity supplied. Because the minimum wage raises the wage that teenagers can earn, it increases the number of teenagers who choose to look for jobs. Studies have found that a higher minimum wage influences which teenagers are employed. When the minimum wage rises, some teenagers who are still attending school choose to drop out and take jobs. These new dropouts displace other teenagers who had already dropped out of school and who now become unemployed.

The minimum wage is a frequent topic of political debate. Advocates of the minimum wage view the policy as one way to raise the income of the working poor. They correctly point out that workers who earn the minimum wage can afford only a meager standard of living. In 1999, for instance, when the minimum wage was \$5.15 per hour, two adults working 40 hours a week for every week of the year at minimum-wage jobs had a total annual income of only \$21,424, which was less than half of the median family income. Many advocates of the minimum wage admit that it has some adverse effects, including unemployment, but they believe that these effects are small and that, all things considered, a higher minimum wage makes the poor better off.

Opponents of the minimum wage contend that it is not the best way to combat poverty. They note that a high minimum wage causes unemployment, encourages teenagers to drop out of school, and prevents some unskilled workers from getting the on-the-job training they need. Moreover, opponents of the minimum wage point out that the minimum wage is a poorly targeted policy. Not all minimum-wage workers are heads of households trying to help their families escape poverty. In fact, fewer than a third of minimum-wage earners are in families with incomes below the poverty line. Many are teenagers from middle-class homes working at part-time jobs for extra spending money.

EVALUATING PRICE CONTROLS

One of the *Ten Principles of Economics* discussed in Chapter 1 is that markets are usually a good way to organize economic activity. This principle explains why





economists usually oppose price ceilings and price floors. To economists, prices are not the outcome of some haphazard process. Prices, they contend, are the result of the millions of business and consumer decisions that lie behind the supply and demand curves. Prices have the crucial job of balancing supply and demand and, thereby, coordinating economic activity. When policymakers set prices by legal decree, they obscure the signals that normally guide the allocation of society's resources.

Another one of the *Ten Principles of Economics* is that governments can sometimes improve market outcomes. Indeed, policymakers are led to control prices because they view the market's outcome as unfair. Price controls are often aimed at helping the poor. For instance, rent-control laws try to make housing affordable for everyone, and minimum-wage laws try to help people escape poverty.

Yet price controls often hurt those they are trying to help. Rent control may keep rents low, but it also discourages landlords from maintaining their buildings and makes housing hard to find. Minimum-wage laws may raise the incomes of some workers, but they also cause other workers to be unemployed.

Helping those in need can be accomplished in ways other than controlling prices. For instance, the government can make housing more affordable by paying a fraction of the rent for poor families. Unlike rent control, such rent subsidies do not reduce the quantity of housing supplied and, therefore, do not lead to housing shortages. Similarly, wage subsidies raise the living standards of the working poor without discouraging firms from hiring them. An example of a wage subsidy is the *earned income tax credit*, a government program that supplements the incomes of low-wage workers.

Although these alternative policies are often better than price controls, they are not perfect. Rent and wage subsidies cost the government money and, therefore, require higher taxes. As we see in the next section, taxation has costs of its own.

QUICK QUIZ: Define *price ceiling* and *price floor*, and give an example of each. Which leads to a shortage? Which leads to a surplus? Why?

TAXES

All governments—from the federal government in Washington, D.C., to the local governments in small towns—use taxes to raise revenue for public projects, such as roads, schools, and national defense. Because taxes are such an important policy instrument, and because they affect our lives in many ways, the study of taxes is a topic to which we return several times throughout this book. In this section we begin our study of how taxes affect the economy.

To set the stage for our analysis, imagine that a local government decides to hold an annual ice-cream celebration—with a parade, fireworks, and speeches by town officials. To raise revenue to pay for the event, it decides to place a \$0.50 tax on the sale of ice-cream cones. When the plan is announced, our two lobbying groups swing into action. The National Organization of Ice Cream Makers claims that its members are struggling to survive in a competitive market, and it argues that *buyers* of ice cream should have to pay the tax. The American Association of Ice Cream Eaters claims that consumers of ice cream are having trouble making ends meet, and it argues that *sellers* of ice cream should pay the tax. The town mayor, hoping to reach a compromise, suggests that half the tax be paid by the buyers and half be paid by the sellers.

To analyze these proposals, we need to address a simple but subtle question: When the government levies a tax on a good, who bears the burden of the tax? The people buying the good? The people selling the good? Or, if buyers and sellers share the tax burden, what determines how the burden is divided? Can the government simply legislate the division of the burden, as the mayor is suggesting, or is the division determined by more fundamental forces in the economy? Economists use the term **tax incidence** to refer to these questions about the distribution of a tax burden. As we will see, we can learn some surprising lessons about tax incidence just by applying the tools of supply and demand.

tax incidence

the study of who bears the burden of taxation

HOW TAXES ON BUYERS AFFECT MARKET OUTCOMES

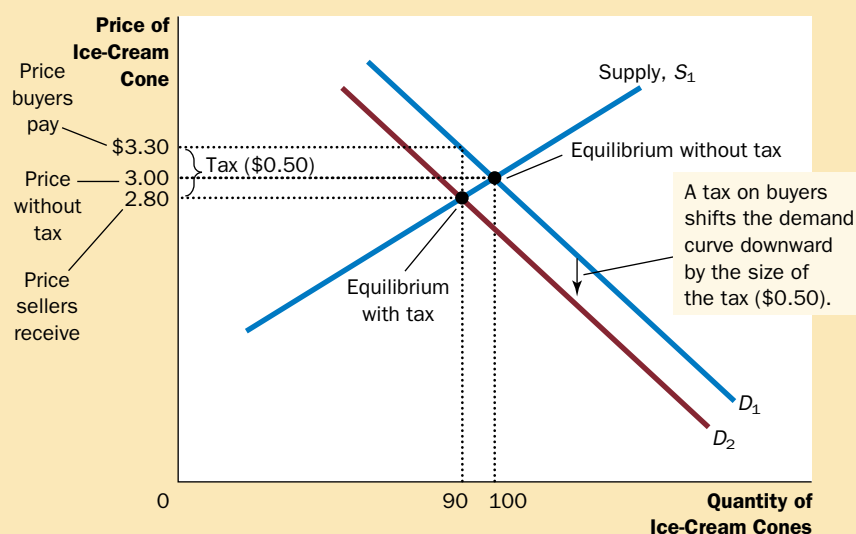
We first consider a tax levied on buyers of a good. Suppose, for instance, that our local government passes a law requiring buyers of ice-cream cones to send \$0.50 to the government for each ice-cream cone they buy. How does this law affect the buyers and sellers of ice cream? To answer this question, we can follow the three steps in Chapter 4 for analyzing supply and demand: (1) We decide whether the law affects the supply curve or demand curve. (2) We decide which way the curve shifts. (3) We examine how the shift affects the equilibrium.

The initial impact of the tax is on the demand for ice cream. The supply curve is not affected because, for any given price of ice cream, sellers have the same incentive to provide ice cream to the market. By contrast, buyers now have to pay a tax to the government (as well as the price to the sellers) whenever they buy ice cream. Thus, the tax shifts the demand curve for ice cream.

The direction of the shift is easy to determine. Because the tax on buyers makes buying ice cream less attractive, buyers demand a smaller quantity of ice cream at every price. As a result, the demand curve shifts to the left (or, equivalently, downward), as shown in Figure 6-6.

Figure 6-6

A TAX ON BUYERS. When a tax of \$0.50 is levied on buyers, the demand curve shifts down by \$0.50 from D_1 to D_2 . The equilibrium quantity falls from 100 to 90 cones. The price that sellers receive falls from \$3.00 to \$2.80. The price that buyers pay (including the tax) rises from \$3.00 to \$3.30. Even though the tax is levied on buyers, buyers and sellers share the burden of the tax.



We can, in this case, be precise about how much the curve shifts. Because of the \$0.50 tax levied on buyers, the effective price to buyers is now \$0.50 higher than the market price. For example, if the market price of a cone happened to be \$2.00, the effective price to buyers would be \$2.50. Because buyers look at their total cost including the tax, they demand a quantity of ice cream as if the market price were \$0.50 higher than it actually is. In other words, to induce buyers to demand any given quantity, the market price must now be \$0.50 lower to make up for the effect of the tax. Thus, the tax shifts the demand curve *downward* from D_1 to D_2 by exactly the size of the tax (\$0.50).

To see the effect of the tax, we compare the old equilibrium and the new equilibrium. You can see in the figure that the equilibrium price of ice cream falls from \$3.00 to \$2.80 and the equilibrium quantity falls from 100 to 90 cones. Because sellers sell less and buyers buy less in the new equilibrium, the tax on ice cream reduces the size of the ice-cream market.

Now let's return to the question of tax incidence: Who pays the tax? Although buyers send the entire tax to the government, buyers and sellers share the burden. Because the market price falls from \$3.00 to \$2.80 when the tax is introduced, sellers receive \$0.20 less for each ice-cream cone than they did without the tax. Thus, the tax makes sellers worse off. Buyers pay sellers a lower price (\$2.80), but the effective price including the tax rises from \$3.00 before the tax to \$3.30 with the tax ($\$2.80 + \$0.50 = \$3.30$). Thus, the tax also makes buyers worse off.

To sum up, the analysis yields two general lessons:

- ◆ Taxes discourage market activity. When a good is taxed, the quantity of the good sold is smaller in the new equilibrium.
- ◆ Buyers and sellers share the burden of taxes. In the new equilibrium, buyers pay more for the good, and sellers receive less.

HOW TAXES ON SELLERS AFFECT MARKET OUTCOMES

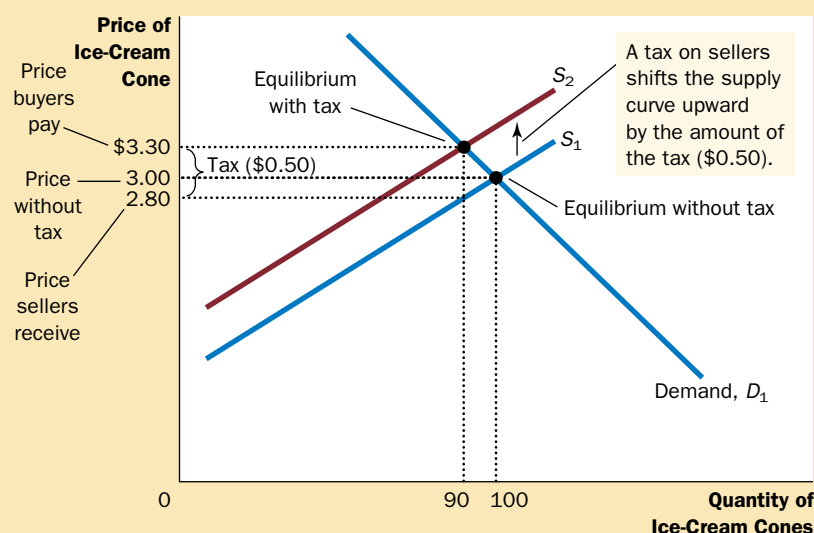
Now consider a tax levied on sellers of a good. Suppose the local government passes a law requiring sellers of ice-cream cones to send \$0.50 to the government for each cone they sell. What are the effects of this law?

In this case, the initial impact of the tax is on the supply of ice cream. Because the tax is not levied on buyers, the quantity of ice cream demanded at any given price is the same, so the demand curve does not change. By contrast, the tax on sellers raises the cost of selling ice cream, and leads sellers to supply a smaller quantity at every price. The supply curve shifts to the left (or, equivalently, upward).

Once again, we can be precise about the magnitude of the shift. For any market price of ice cream, the effective price to sellers—the amount they get to keep after paying the tax—is \$0.50 lower. For example, if the market price of a cone happened to be \$2.00, the effective price received by sellers would be \$1.50. Whatever the market price, sellers will supply a quantity of ice cream as if the price were \$0.50 lower than it is. Put differently, to induce sellers to supply any given quantity, the market price must now be \$0.50 higher to compensate for the effect of the tax. Thus, as shown in Figure 6-7, the supply curve shifts *upward* from S_1 to S_2 by exactly the size of the tax (\$0.50).

When the market moves from the old to the new equilibrium, the equilibrium price of ice cream rises from \$3.00 to \$3.30, and the equilibrium quantity falls from

Figure 6-7



A TAX ON SELLERS. When a tax of \$0.50 is levied on sellers, the supply curve shifts up by \$0.50 from S_1 to S_2 . The equilibrium quantity falls from 100 to 90 cones. The price that buyers pay rises from \$3.00 to \$3.30. The price that sellers receive (after paying the tax) falls from \$3.00 to \$2.80. Even though the tax is levied on sellers, buyers and sellers share the burden of the tax.

100 to 90 cones. Once again, the tax reduces the size of the ice-cream market. And once again, buyers and sellers share the burden of the tax. Because the market price rises, buyers pay \$0.30 more for each cone than they did before the tax was enacted. Sellers receive a higher price than they did without the tax, but the effective price (after paying the tax) falls from \$3.00 to \$2.80.

Comparing Figures 6-6 and 6-7 leads to a surprising conclusion: *Taxes on buyers and taxes on sellers are equivalent.* In both cases, the tax places a wedge between the price that buyers pay and the price that sellers receive. The wedge between the buyers' price and the sellers' price is the same, regardless of whether the tax is levied on buyers or sellers. In either case, the wedge shifts the relative position of the supply and demand curves. In the new equilibrium, buyers and sellers share the burden of the tax. The only difference between taxes on buyers and taxes on sellers is who sends the money to the government.

The equivalence of these two taxes is perhaps easier to understand if we imagine that the government collects the \$0.50 ice-cream tax in a bowl on the counter of each ice-cream store. When the government levies the tax on buyers, the buyer is required to place \$0.50 in the bowl every time a cone is bought. When the government levies the tax on sellers, the seller is required to place \$0.50 in the bowl after the sale of each cone. Whether the \$0.50 goes directly from the buyer's pocket into the bowl, or indirectly from the buyer's pocket into the seller's hand and then into the bowl, does not matter. Once the market reaches its new equilibrium, buyers and sellers share the burden, regardless of how the tax is levied.

CASE STUDY CAN CONGRESS DISTRIBUTE THE BURDEN OF A PAYROLL TAX?

If you have ever received a paycheck, you probably noticed that taxes were deducted from the amount you earned. One of these taxes is called FICA, an

acronym for the Federal Insurance Contribution Act. The federal government uses the revenue from the FICA tax to pay for Social Security and Medicare, the income support and health care programs for the elderly. FICA is an example of a *payroll tax*, which is a tax on the wages that firms pay their workers. In 1999, the total FICA tax for the typical worker was 15.3 percent of earnings.

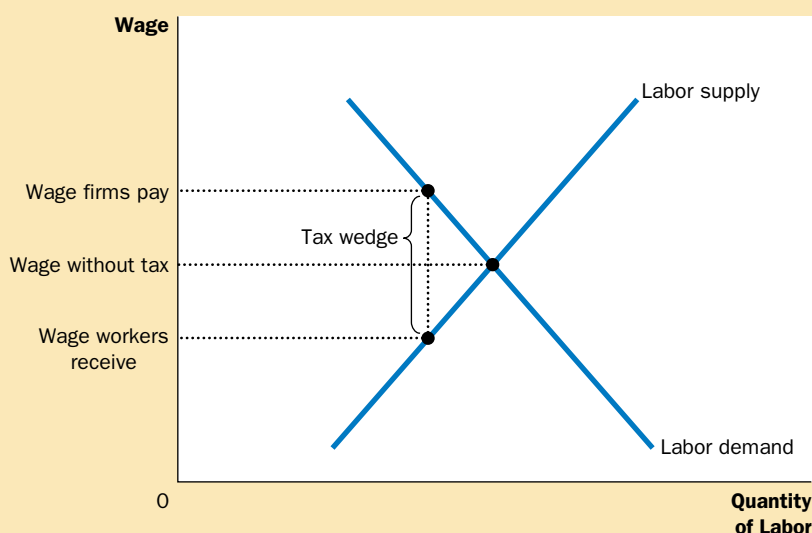
Who do you think bears the burden of this payroll tax—firms or workers? When Congress passed this legislation, it attempted to mandate a division of the tax burden. According to the law, half of the tax is paid by firms, and half is paid by workers. That is, half of the tax is paid out of firm revenue, and half is deducted from workers' paychecks. The amount that shows up as a deduction on your pay stub is the worker contribution.

Our analysis of tax incidence, however, shows that lawmakers cannot so easily distribute the burden of a tax. To illustrate, we can analyze a payroll tax as merely a tax on a good, where the good is labor and the price is the wage. The key feature of the payroll tax is that it places a wedge between the wage that firms pay and the wage that workers receive. Figure 6-8 shows the outcome. When a payroll tax is enacted, the wage received by workers falls, and the wage paid by firms rises. In the end, workers and firms share the burden of the tax, much as the legislation requires. Yet this division of the tax burden between workers and firms has nothing to do with the legislated division: The division of the burden in Figure 6-8 is not necessarily fifty-fifty, and the same outcome would prevail if the law levied the entire tax on workers or if it levied the entire tax on firms.

This example shows that the most basic lesson of tax incidence is often overlooked in public debate. Lawmakers can decide whether a tax comes from the buyer's pocket or from the seller's, but they cannot legislate the true burden of a tax. Rather, tax incidence depends on the forces of supply and demand.

Figure 6-8

A PAYROLL TAX. A payroll tax places a wedge between the wage that workers receive and the wage that firms pay. Comparing wages with and without the tax, you can see that workers and firms share the tax burden. This division of the tax burden between workers and firms does not depend on whether the government levies the tax on workers, levies the tax on firms, or divides the tax equally between the two groups.

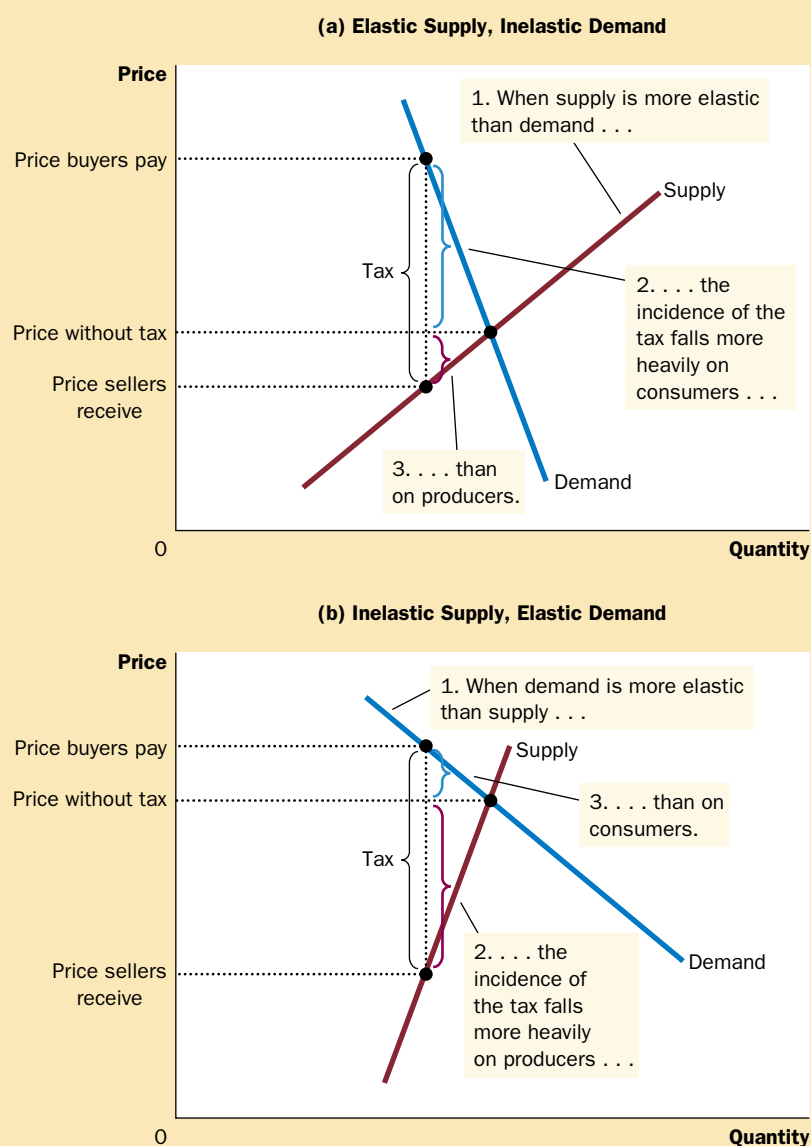


ELASTICITY AND TAX INCIDENCE

When a good is taxed, buyers and sellers of the good share the burden of the tax. But how exactly is the tax burden divided? Only rarely will it be shared equally. To see how the burden is divided, consider the impact of taxation in the two markets in Figure 6-9. In both cases, the figure shows the initial demand curve, the initial supply curve, and a tax that drives a wedge between the amount paid by buyers and the amount received by sellers. (Not drawn in either panel of the figure is the new supply or demand curve. Which curve shifts depends on whether the tax is levied on buyers or sellers. As we have seen, this is irrelevant for the incidence of

Figure 6-9

HOW THE BURDEN OF A TAX IS DIVIDED. In panel (a), the supply curve is elastic, and the demand curve is inelastic. In this case, the price received by sellers falls only slightly, while the price paid by buyers rises substantially. Thus, buyers bear most of the burden of the tax. In panel (b), the supply curve is inelastic, and the demand curve is elastic. In this case, the price received by sellers falls substantially, while the price paid by buyers rises only slightly. Thus, sellers bear most of the burden of the tax.



the tax.) The difference in the two panels is the relative elasticity of supply and demand.

Panel (a) of Figure 6-9 shows a tax in a market with very elastic supply and relatively inelastic demand. That is, sellers are very responsive to the price of the good, whereas buyers are not very responsive. When a tax is imposed on a market with these elasticities, the price received by sellers does not fall much, so sellers bear only a small burden. By contrast, the price paid by buyers rises substantially, indicating that buyers bear most of the burden of the tax.

Panel (b) of Figure 6-9 shows a tax in a market with relatively inelastic supply and very elastic demand. In this case, sellers are not very responsive to the price, while buyers are very responsive. The figure shows that when a tax is imposed, the price paid by buyers does not rise much, while the price received by sellers falls substantially. Thus, sellers bear most of the burden of the tax.

The two panels of Figure 6-9 show a general lesson about how the burden of a tax is divided: *A tax burden falls more heavily on the side of the market that is less elastic.* Why is this true? In essence, the elasticity measures the willingness of buyers or sellers to leave the market when conditions become unfavorable. A small elasticity of demand means that buyers do not have good alternatives to consuming this particular good. A small elasticity of supply means that sellers do not have good alternatives to producing this particular good. When the good is taxed, the side of the market with fewer good alternatives cannot easily leave the market and must, therefore, bear more of the burden of the tax.

We can apply this logic to the payroll tax, which was discussed in the previous case study. Most labor economists believe that the supply of labor is much less elastic than the demand. This means that workers, rather than firms, bear most of the burden of the payroll tax. In other words, the distribution of the tax burden is not at all close to the fifty-fifty split that lawmakers intended.



"If this boat were any more expensive, we would be playing golf."

CASE STUDY WHO PAYS THE LUXURY TAX?

In 1990, Congress adopted a new luxury tax on items such as yachts, private airplanes, furs, jewelry, and expensive cars. The goal of the tax was to raise revenue from those who could most easily afford to pay. Because only the rich could afford to buy such extravagances, taxing luxuries seemed a logical way of taxing the rich.

Yet, when the forces of supply and demand took over, the outcome was quite different from what Congress intended. Consider, for example, the market for yachts. The demand for yachts is quite elastic. A millionaire can easily not buy a yacht; she can use the money to buy a bigger house, take a European vacation, or leave a larger bequest to her heirs. By contrast, the supply of yachts is relatively inelastic, at least in the short run. Yacht factories are not easily converted to alternative uses, and workers who build yachts are not eager to change careers in response to changing market conditions.

Our analysis makes a clear prediction in this case. With elastic demand and inelastic supply, the burden of a tax falls largely on the suppliers. That is, a tax on yachts places a burden largely on the firms and workers who build yachts because they end up getting a lower price for their product. The workers, however, are not wealthy. Thus, the burden of a luxury tax falls more on the middle class than on the rich.

The mistaken assumptions about the incidence of the luxury tax quickly became apparent after the tax went into effect. Suppliers of luxuries made their congressional representatives well aware of the economic hardship they experienced, and Congress repealed most of the luxury tax in 1993.

QUICK QUIZ: In a supply-and-demand diagram, show how a tax on car buyers of \$1,000 per car affects the quantity of cars sold and the price of cars. In another diagram, show how a tax on car sellers of \$1,000 per car affects the quantity of cars sold and the price of cars. In both of your diagrams, show the change in the price paid by car buyers and the change in price received by car sellers.

CONCLUSION

The economy is governed by two kinds of laws: the laws of supply and demand and the laws enacted by governments. In this chapter we have begun to see how these laws interact. Price controls and taxes are common in various markets in the economy, and their effects are frequently debated in the press and among policy-makers. Even a little bit of economic knowledge can go a long way toward understanding and evaluating these policies.

In subsequent chapters we will analyze many government policies in greater detail. We will examine the effects of taxation more fully, and we will consider a broader range of policies than we considered here. Yet the basic lessons of this chapter will not change: When analyzing government policies, supply and demand are the first and most useful tools of analysis.

Summary

- ◆ A price ceiling is a legal maximum on the price of a good or service. An example is rent control. If the price ceiling is below the equilibrium price, the quantity demanded exceeds the quantity supplied. Because of the resulting shortage, sellers must in some way ration the good or service among buyers.
- ◆ A price floor is a legal minimum on the price of a good or service. An example is the minimum wage. If the price floor is above the equilibrium price, the quantity supplied exceeds the quantity demanded. Because of the resulting surplus, buyers' demands for the good or service must in some way be rationed among sellers.
- ◆ When the government levies a tax on a good, the equilibrium quantity of the good falls. That is, a tax on a market shrinks the size of the market.
- ◆ A tax on a good places a wedge between the price paid by buyers and the price received by sellers. When the market moves to the new equilibrium, buyers pay more for the good and sellers receive less for it. In this sense, buyers and sellers share the tax burden. The incidence of a tax does not depend on whether the tax is levied on buyers or sellers.
- ◆ The incidence of a tax depends on the price elasticities of supply and demand. The burden tends to fall on the side of the market that is less elastic because that side of the market can respond less easily to the tax by changing the quantity bought or sold.

Key Concepts

price ceiling, p. 118

price floor, p. 118

tax incidence, p. 129

Questions for Review

1. Give an example of a price ceiling and an example of a price floor.
2. Which causes a shortage of a good—a price ceiling or a price floor? Which causes a surplus?
3. What mechanisms allocate resources when the price of a good is not allowed to bring supply and demand into equilibrium?
4. Explain why economists usually oppose controls on prices.
5. What is the difference between a tax paid by buyers and a tax paid by sellers?
6. How does a tax on a good affect the price paid by buyers, the price received by sellers, and the quantity sold?
7. What determines how the burden of a tax is divided between buyers and sellers? Why?

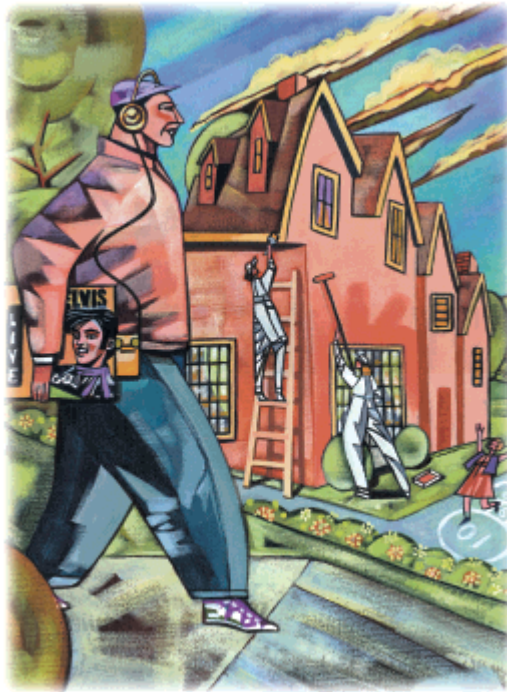
Problems and Applications

1. Lovers of classical music persuade Congress to impose a price ceiling of \$40 per ticket. Does this policy get more or fewer people to attend classical music concerts?
2. The government has decided that the free-market price of cheese is too low.
 - a. Suppose the government imposes a binding price floor in the cheese market. Use a supply-and-demand diagram to show the effect of this policy on the price of cheese and the quantity of cheese sold. Is there a shortage or surplus of cheese?
 - b. Farmers complain that the price floor has reduced their total revenue. Is this possible? Explain.
 - c. In response to farmers' complaints, the government agrees to purchase all of the surplus cheese at the price floor. Compared to the basic price floor, who benefits from this new policy? Who loses?
3. A recent study found that the demand and supply schedules for Frisbees are as follows:
 - a. What are the equilibrium price and quantity of Frisbees?
 - b. Frisbee manufacturers persuade the government that Frisbee production improves scientists' understanding of aerodynamics and thus is important for national security. A concerned Congress votes to impose a price floor \$2 above the equilibrium price. What is the new market price? How many Frisbees are sold?
 - c. Irate college students march on Washington and demand a reduction in the price of Frisbees. An even more concerned Congress votes to repeal the price floor and impose a price ceiling \$1 below the former price floor. What is the new market price? How many Frisbees are sold?
4. Suppose the federal government requires beer drinkers to pay a \$2 tax on each case of beer purchased. (In fact, both the federal and state governments impose beer taxes of some sort.)
 - a. Draw a supply-and-demand diagram of the market for beer without the tax. Show the price paid by consumers, the price received by producers, and the quantity of beer sold. What is the difference between the price paid by consumers and the price received by producers?
 - b. Now draw a supply-and-demand diagram for the beer market with the tax. Show the price paid by consumers, the price received by producers, and

PRICE PER FRISBEE	QUANTITY DEMANDED	QUANTITY SUPPLIED
\$11	1 million	15 million
10	2	12
9	4	9
8	6	6
7	8	3
6	10	1

- the quantity of beer sold. What is the difference between the price paid by consumers and the price received by producers? Has the quantity of beer sold increased or decreased?
5. A senator wants to raise tax revenue and make workers better off. A staff member proposes raising the payroll tax paid by firms and using part of the extra revenue to reduce the payroll tax paid by workers. Would this accomplish the senator's goal?
 6. If the government places a \$500 tax on luxury cars, will the price paid by consumers rise by more than \$500, less than \$500, or exactly \$500? Explain.
 7. Congress and the president decide that the United States should reduce air pollution by reducing its use of gasoline. They impose a \$0.50 tax for each gallon of gasoline sold.
 - a. Should they impose this tax on producers or consumers? Explain carefully using a supply-and-demand diagram.
 - b. If the demand for gasoline were more elastic, would this tax be more effective or less effective in reducing the quantity of gasoline consumed? Explain with both words and a diagram.
 - c. Are consumers of gasoline helped or hurt by this tax? Why?
 - d. Are workers in the oil industry helped or hurt by this tax? Why?
 8. A case study in this chapter discusses the federal minimum-wage law.
 - a. Suppose the minimum wage is above the equilibrium wage in the market for unskilled labor. Using a supply-and-demand diagram of the market for unskilled labor, show the market wage, the number of workers who are employed, and the number of workers who are unemployed. Also show the total wage payments to unskilled workers.
 - b. Now suppose the secretary of labor proposes an increase in the minimum wage. What effect would this increase have on employment? Does the change in employment depend on the elasticity of demand, the elasticity of supply, both elasticities, or neither?
 - c. What effect would this increase in the minimum wage have on unemployment? Does the change in unemployment depend on the elasticity of demand, the elasticity of supply, both elasticities, or neither?
 - d. If the demand for unskilled labor were inelastic, would the proposed increase in the minimum wage raise or lower total wage payments to unskilled workers? Would your answer change if the demand for unskilled labor were elastic?
 9. Consider the following policies, each of which is aimed at reducing violent crime by reducing the use of guns. Illustrate each of these proposed policies in a supply-and-demand diagram of the gun market.
 - a. a tax on gun buyers
 - b. a tax on gun sellers
 - c. a price floor on guns
 - d. a tax on ammunition
 10. The U.S. government administers two programs that affect the market for cigarettes. Media campaigns and labeling requirements are aimed at making the public aware of the dangers of cigarette smoking. At the same time, the Department of Agriculture maintains a price support program for tobacco farmers, which raises the price of tobacco above the equilibrium price.
 - a. How do these two programs affect cigarette consumption? Use a graph of the cigarette market in your answer.
 - b. What is the combined effect of these two programs on the price of cigarettes?
 - c. Cigarettes are also heavily taxed. What effect does this tax have on cigarette consumption?
 11. A subsidy is the opposite of a tax. With a \$0.50 tax on the buyers of ice-cream cones, the government collects \$0.50 for each cone purchased; with a \$0.50 subsidy for the buyers of ice-cream cones, the government pays buyers \$0.50 for each cone purchased.
 - a. Show the effect of a \$0.50 per cone subsidy on the demand curve for ice-cream cones, the effective price paid by consumers, the effective price received by sellers, and the quantity of cones sold.
 - b. Do consumers gain or lose from this policy? Do producers gain or lose? Does the government gain or lose?

7



CONSUMERS, PRODUCERS, AND THE EFFICIENCY OF MARKETS

When consumers go to grocery stores to buy their turkeys for Thanksgiving dinner, they may be disappointed that the price of turkey is as high as it is. At the same time, when farmers bring to market the turkeys they have raised, they wish the price of turkey were even higher. These views are not surprising: Buyers always want to pay less, and sellers always want to get paid more. But is there a “right price” for turkey from the standpoint of society as a whole?

In previous chapters we saw how, in market economies, the forces of supply and demand determine the prices of goods and services and the quantities sold. So far, however, we have described the way markets allocate scarce resources without directly addressing the question of whether these market allocations are desirable. In other words, our analysis has been *positive* (what is) rather than *normative* (what

IN THIS CHAPTER YOU WILL . . .

Examine the link between buyers' willingness to pay for a good and the demand curve

Learn how to define and measure consumer surplus

Examine the link between sellers' costs of producing a good and the supply curve

Learn how to define and measure producer surplus

See that the equilibrium of supply and demand maximizes total surplus in a market

welfare economics
*the study of how the allocation of
resources affects economic well-being*



should be). We know that the price of turkey adjusts to ensure that the quantity of turkey supplied equals the quantity of turkey demanded. But, at this equilibrium, is the quantity of turkey produced and consumed too small, too large, or just right?

In this chapter we take up the topic of **welfare economics**, the study of how the allocation of resources affects economic well-being. We begin by examining the benefits that buyers and sellers receive from taking part in a market. We then examine how society can make these benefits as large as possible. This analysis leads to a profound conclusion: The equilibrium of supply and demand in a market maximizes the total benefits received by buyers and sellers.

As you may recall from Chapter 1, one of the *Ten Principles of Economics* is that markets are usually a good way to organize economic activity. The study of welfare economics explains this principle more fully. It also answers our question about the right price of turkey: The price that balances the supply and demand for turkey is, in a particular sense, the best one because it maximizes the total welfare of turkey consumers and turkey producers.

CONSUMER SURPLUS

We begin our study of welfare economics by looking at the benefits buyers receive from participating in a market.

WILLINGNESS TO PAY

Imagine that you own a mint-condition recording of Elvis Presley’s first album. Because you are not an Elvis Presley fan, you decide to sell it. One way to do so is to hold an auction.

Four Elvis fans show up for your auction: John, Paul, George, and Ringo. Each of them would like to own the album, but there is a limit to the amount that each is willing to pay for it. Table 7-1 shows the maximum price that each of the four possible buyers would pay. Each buyer’s maximum is called his **willingness to pay**, and it measures how much that buyer values the good. Each buyer would be eager to buy the album at a price less than his willingness to pay, would refuse to

willingness to pay
*the maximum amount that a buyer
will pay for a good*

Table 7-1

FOUR POSSIBLE BUYERS’
WILLINGNESS TO PAY

BUYER	WILLINGNESS TO PAY
John	\$100
Paul	80
George	70
Ringo	50

buy the album at a price more than his willingness to pay, and would be indifferent about buying the album at a price exactly equal to his willingness to pay.

To sell your album, you begin the bidding at a low price, say \$10. Because all four buyers are willing to pay much more, the price rises quickly. The bidding stops when John bids \$80 (or slightly more). At this point, Paul, George, and Ringo have dropped out of the bidding, because they are unwilling to bid any more than \$80. John pays you \$80 and gets the album. Note that the album has gone to the buyer who values the album most highly.

What benefit does John receive from buying the Elvis Presley album? In a sense, John has found a real bargain: He is willing to pay \$100 for the album but pays only \$80 for it. We say that John receives *consumer surplus* of \$20. **Consumer surplus** is the amount a buyer is willing to pay for a good minus the amount the buyer actually pays for it.

consumer surplus

a buyer's willingness to pay minus the amount the buyer actually pays

Consumer surplus measures the benefit to buyers of participating in a market. In this example, John receives a \$20 benefit from participating in the auction because he pays only \$80 for a good he values at \$100. Paul, George, and Ringo get no consumer surplus from participating in the auction, because they left without the album and without paying anything.

Now consider a somewhat different example. Suppose that you had two identical Elvis Presley albums to sell. Again, you auction them off to the four possible buyers. To keep things simple, we assume that both albums are to be sold for the same price and that no buyer is interested in buying more than one album. Therefore, the price rises until two buyers are left.

In this case, the bidding stops when John and Paul bid \$70 (or slightly higher). At this price, John and Paul are each happy to buy an album, and George and Ringo are not willing to bid any higher. John and Paul each receive consumer surplus equal to his willingness to pay minus the price. John's consumer surplus is \$30, and Paul's is \$10. John's consumer surplus is higher now than it was previously, because he gets the same album but pays less for it. The total consumer surplus in the market is \$40.

USING THE DEMAND CURVE TO MEASURE CONSUMER SURPLUS

Consumer surplus is closely related to the demand curve for a product. To see how they are related, let's continue our example and consider the demand curve for this rare Elvis Presley album.

We begin by using the willingness to pay of the four possible buyers to find the demand schedule for the album. Table 7-2 shows the demand schedule that corresponds to Table 7-1. If the price is above \$100, the quantity demanded in the market is 0, because no buyer is willing to pay that much. If the price is between \$80 and \$100, the quantity demanded is 1, because only John is willing to pay such a high price. If the price is between \$70 and \$80, the quantity demanded is 2, because both John and Paul are willing to pay the price. We can continue this analysis for other prices as well. In this way, the demand schedule is derived from the willingness to pay of the four possible buyers.

Figure 7-1 graphs the demand curve that corresponds to this demand schedule. Note the relationship between the height of the demand curve and the buyers' willingness to pay. At any quantity, the price given by the demand curve shows

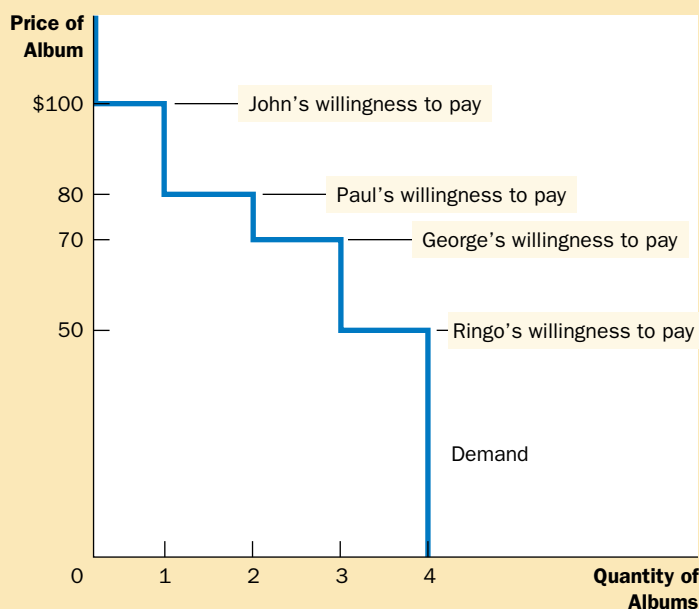
Table 7-2

THE DEMAND SCHEDULE FOR THE BUYERS IN TABLE 7-1

PRICE	BUYERS	QUANTITY DEMANDED
More than \$100	None	0
\$80 to \$100	John	1
\$70 to \$80	John, Paul	2
\$50 to \$70	John, Paul, George	3
\$50 or less	John, Paul, George, Ringo	4

Figure 7-1

THE DEMAND CURVE. This figure graphs the demand curve from the demand schedule in Table 7-2. Note that the height of the demand curve reflects buyers' willingness to pay.



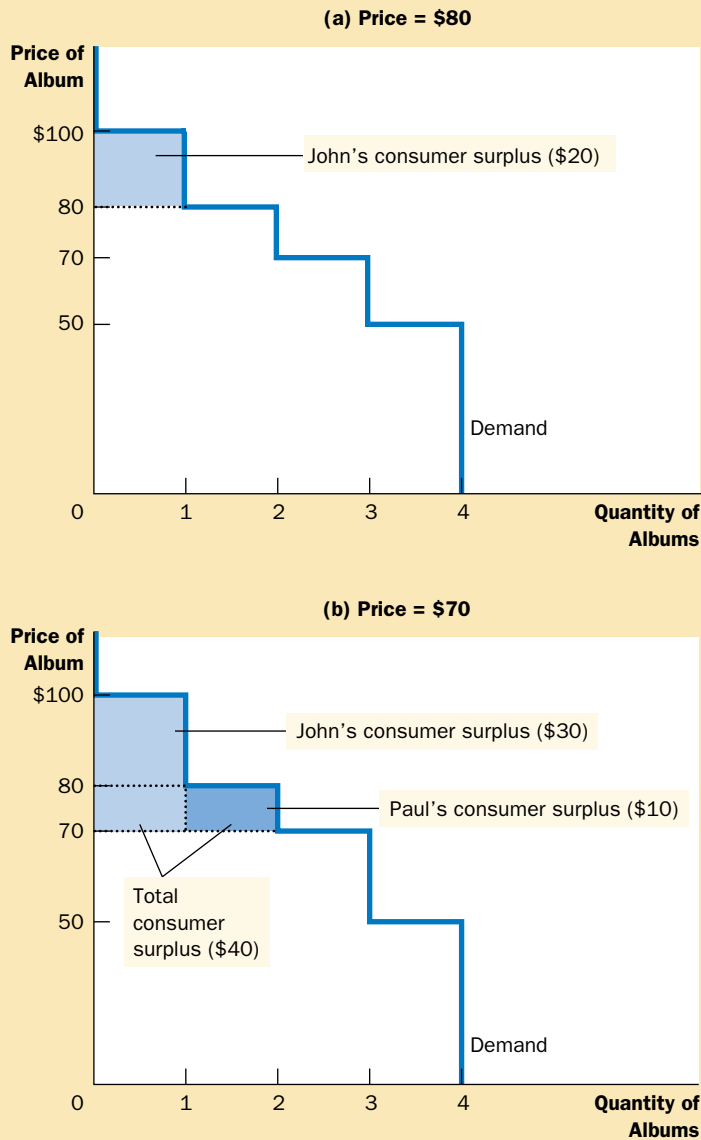
the willingness to pay of the *marginal buyer*, the buyer who would leave the market first if the price were any higher. At a quantity of 4 albums, for instance, the demand curve has a height of \$50, the price that Ringo (the marginal buyer) is willing to pay for an album. At a quantity of 3 albums, the demand curve has a height of \$70, the price that George (who is now the marginal buyer) is willing to pay.

Because the demand curve reflects buyers' willingness to pay, we can also use it to measure consumer surplus. Figure 7-2 uses the demand curve to compute consumer surplus in our example. In panel (a), the price is \$80 (or slightly above), and the quantity demanded is 1. Note that the area above the price and below the demand curve equals \$20. This amount is exactly the consumer surplus we computed earlier when only 1 album is sold.

Panel (b) of Figure 7-2 shows consumer surplus when the price is \$70 (or slightly above). In this case, the area above the price and below the demand curve

Figure 7-2

MEASURING CONSUMER SURPLUS WITH THE DEMAND CURVE. In panel (a), the price of the good is \$80, and the consumer surplus is \$20. In panel (b), the price of the good is \$70, and the consumer surplus is \$40.



equals the total area of the two rectangles: John's consumer surplus at this price is \$30 and Paul's is \$10. This area equals a total of \$40. Once again, this amount is the consumer surplus we computed earlier.

The lesson from this example holds for all demand curves: *The area below the demand curve and above the price measures the consumer surplus in a market.* The reason is that the height of the demand curve measures the value buyers place on the good, as measured by their willingness to pay for it. The difference between this willingness to pay and the market price is each buyer's consumer surplus. Thus, the total area below the demand curve and above the price is the sum of the consumer surplus of all buyers in the market for a good or service.

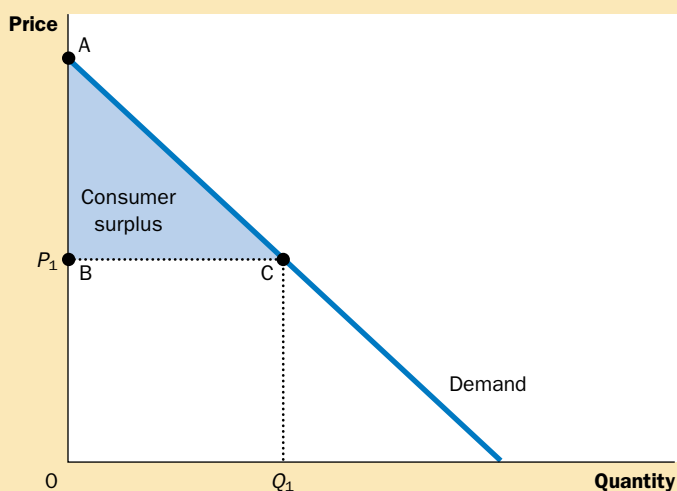
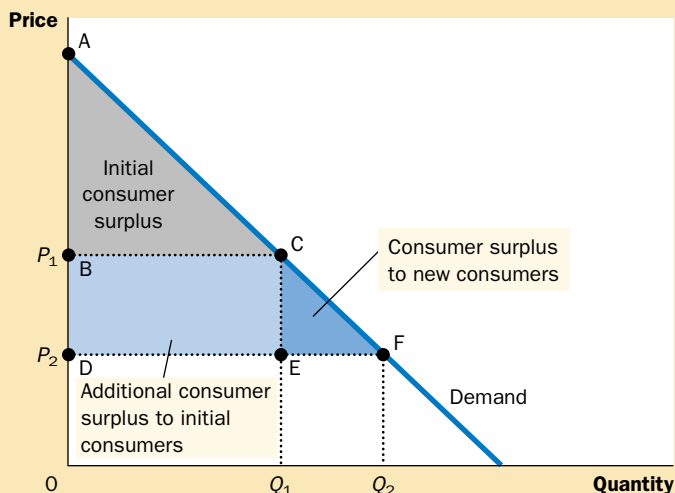
HOW A LOWER PRICE RAISES CONSUMER SURPLUS

Because buyers always want to pay less for the goods they buy, a lower price makes buyers of a good better off. But how much does buyers' well-being rise in response to a lower price? We can use the concept of consumer surplus to answer this question precisely.

Figure 7-3 shows a typical downward-sloping demand curve. Although this demand curve appears somewhat different in shape from the steplike demand curves in our previous two figures, the ideas we have just developed apply nonetheless: Consumer surplus is the area above the price and below the demand curve. In panel (a), consumer surplus at a price of P_1 is the area of triangle ABC.

Figure 7-3

HOW THE PRICE AFFECTS CONSUMER SURPLUS. In panel (a), the price is P_1 , the quantity demanded is Q_1 , and consumer surplus equals the area of the triangle ABC. When the price falls from P_1 to P_2 , as in panel (b), the quantity demanded rises from Q_1 to Q_2 , and the consumer surplus rises to the area of the triangle ADF. The increase in consumer surplus (area BCFD) occurs in part because existing consumers now pay less (area BCED) and in part because new consumers enter the market at the lower price (area CEF).

(a) Consumer Surplus at Price P_1 **(b) Consumer Surplus at Price P_2** 

Now suppose that the price falls from P_1 to P_2 , as shown in panel (b). The consumer surplus now equals area ADF. The increase in consumer surplus attributable to the lower price is the area BCFD.

This increase in consumer surplus is composed of two parts. First, those buyers who were already buying Q_1 of the good at the higher price P_1 are better off because they now pay less. The increase in consumer surplus of existing buyers is the reduction in the amount they pay; it equals the area of the rectangle BCED. Second, some new buyers enter the market because they are now willing to buy the good at the lower price. As a result, the quantity demanded in the market increases from Q_1 to Q_2 . The consumer surplus these newcomers receive is the area of the triangle CEF.

WHAT DOES CONSUMER SURPLUS MEASURE?

Our goal in developing the concept of consumer surplus is to make normative judgments about the desirability of market outcomes. Now that you have seen what consumer surplus is, let's consider whether it is a good measure of economic well-being.

Imagine that you are a policymaker trying to design a good economic system. Would you care about the amount of consumer surplus? Consumer surplus, the amount that buyers are willing to pay for a good minus the amount they actually pay for it, measures the benefit that buyers receive from a good *as the buyers themselves perceive it*. Thus, consumer surplus is a good measure of economic well-being if policymakers want to respect the preferences of buyers.

In some circumstances, policymakers might choose not to care about consumer surplus because they do not respect the preferences that drive buyer behavior. For example, drug addicts are willing to pay a high price for heroin. Yet we would not say that addicts get a large benefit from being able to buy heroin at a low price (even though addicts might say they do). From the standpoint of society, willingness to pay in this instance is not a good measure of the buyers' benefit, and consumer surplus is not a good measure of economic well-being, because addicts are not looking after their own best interests.

In most markets, however, consumer surplus does reflect economic well-being. Economists normally presume that buyers are rational when they make decisions and that their preferences should be respected. In this case, consumers are the best judges of how much benefit they receive from the goods they buy.

QUICK QUIZ: Draw a demand curve for turkey. In your diagram, show a price of turkey and the consumer surplus that results from that price. Explain in words what this consumer surplus measures.

PRODUCER SURPLUS

We now turn to the other side of the market and consider the benefits sellers receive from participating in a market. As you will see, our analysis of sellers' welfare is similar to our analysis of buyers' welfare.

COST AND THE WILLINGNESS TO SELL

cost
the value of everything a seller must give up to produce a good

Imagine now that you are a homeowner, and you need to get your house painted. You turn to four sellers of painting services: Mary, Frida, Georgia, and Grandma. Each painter is willing to do the work for you if the price is right. You decide to take bids from the four painters and auction off the job to the painter who will do the work for the lowest price.

Each painter is willing to take the job if the price she would receive exceeds her cost of doing the work. Here the term **cost** should be interpreted as the painters’ opportunity cost: It includes the painters’ out-of-pocket expenses (for paint, brushes, and so on) as well as the value that the painters place on their own time. Table 7-3 shows each painter’s cost. Because a painter’s cost is the lowest price she would accept for her work, cost is a measure of her willingness to sell her services. Each painter would be eager to sell her services at a price greater than her cost, would refuse to sell her services at a price less than her cost, and would be indifferent about selling her services at a price exactly equal to her cost.

When you take bids from the painters, the price might start off high, but it quickly falls as the painters compete for the job. Once Grandma has bid \$600 (or slightly less), she is the sole remaining bidder. Grandma is happy to do the job for this price, because her cost is only \$500. Mary, Frida, and Georgia are unwilling to do the job for less than \$600. Note that the job goes to the painter who can do the work at the lowest cost.

producer surplus
the amount a seller is paid for a good minus the seller’s cost

What benefit does Grandma receive from getting the job? Because she is willing to do the work for \$500 but gets \$600 for doing it, we say that she receives *producer surplus* of \$100. **Producer surplus** is the amount a seller is paid minus the cost of production. Producer surplus measures the benefit to sellers of participating in a market.

Now consider a somewhat different example. Suppose that you have two houses that need painting. Again, you auction off the jobs to the four painters. To keep things simple, let’s assume that no painter is able to paint both houses and that you will pay the same amount to paint each house. Therefore, the price falls until two painters are left.

In this case, the bidding stops when Georgia and Grandma each offer to do the job for a price of \$800 (or slightly less). At this price, Georgia and Grandma are willing to do the work, and Mary and Frida are not willing to bid a lower price. At a price of \$800, Grandma receives producer surplus of \$300, and Georgia receives producer surplus of \$200. The total producer surplus in the market is \$500.

Table 7-3
THE COSTS OF FOUR POSSIBLE
SELLERS

SELLER	COST
Mary	\$900
Frida	800
Georgia	600
Grandma	500

USING THE SUPPLY CURVE TO MEASURE PRODUCER SURPLUS

Just as consumer surplus is closely related to the demand curve, producer surplus is closely related to the supply curve. To see how, let's continue our example.

We begin by using the costs of the four painters to find the supply schedule for painting services. Table 7-4 shows the supply schedule that corresponds to the costs in Table 7-3. If the price is below \$500, none of the four painters is willing to do the job, so the quantity supplied is zero. If the price is between \$500 and \$600, only Grandma is willing to do the job, so the quantity supplied is 1. If the price is between \$600 and \$800, Grandma and Georgia are willing to do the job, so the quantity supplied is 2, and so on. Thus, the supply schedule is derived from the costs of the four painters.

Figure 7-4 graphs the supply curve that corresponds to this supply schedule. Note that the height of the supply curve is related to the sellers' costs. At any quantity, the price given by the supply curve shows the cost of the *marginal seller*, the

Table 7-4

THE SUPPLY SCHEDULE FOR THE SELLERS IN TABLE 7-3

PRICE	SELLERS	QUANTITY SUPPLIED
\$900 or more	Mary, Frida, Georgia, Grandma	4
\$800 to \$900	Frida, Georgia, Grandma	3
\$600 to \$800	Georgia, Grandma	2
\$500 to \$600	Grandma	1
Less than \$500	None	0

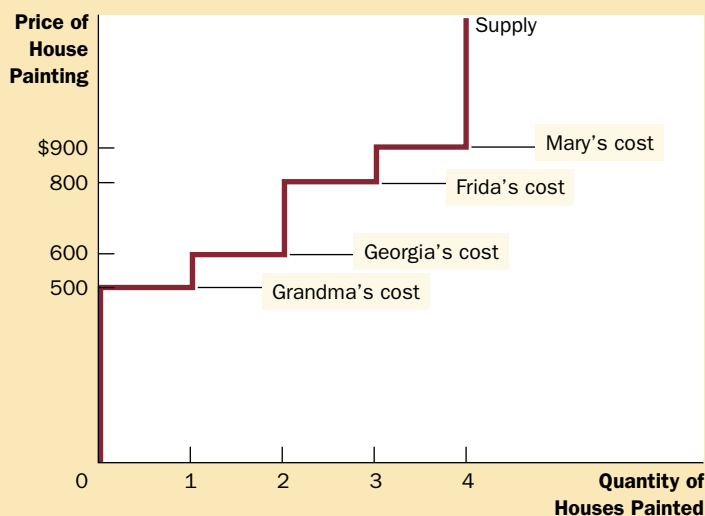


Figure 7-4

THE SUPPLY CURVE. This figure graphs the supply curve from the supply schedule in Table 7-4. Note that the height of the supply curve reflects sellers' costs.

seller who would leave the market first if the price were any lower. At a quantity of 4 houses, for instance, the supply curve has a height of \$900, the cost that Mary (the marginal seller) incurs to provide her painting services. At a quantity of 3 houses, the supply curve has a height of \$800, the cost that Frida (who is now the marginal seller) incurs.

Because the supply curve reflects sellers' costs, we can use it to measure producer surplus. Figure 7-5 uses the supply curve to compute producer surplus in our example. In panel (a), we assume that the price is \$600. In this case, the quantity supplied is 1. Note that the area below the price and above the supply curve equals \$100. This amount is exactly the producer surplus we computed earlier for Grandma.

Panel (b) of Figure 7-5 shows producer surplus at a price of \$800. In this case, the area below the price and above the supply curve equals the total area of the two rectangles. This area equals \$500, the producer surplus we computed earlier for Georgia and Grandma when two houses needed painting.

The lesson from this example applies to all supply curves: *The area below the price and above the supply curve measures the producer surplus in a market.* The logic is straightforward: The height of the supply curve measures sellers' costs, and the difference between the price and the cost of production is each seller's producer surplus. Thus, the total area is the sum of the producer surplus of all sellers.

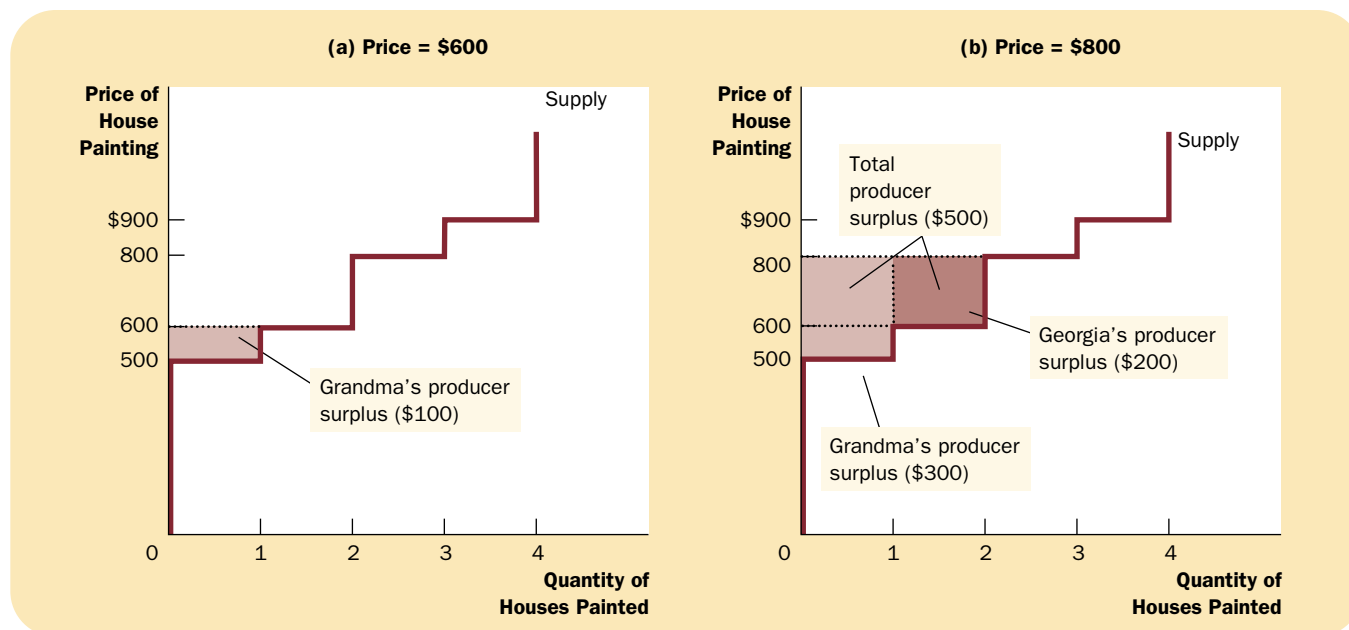


Figure 7-5

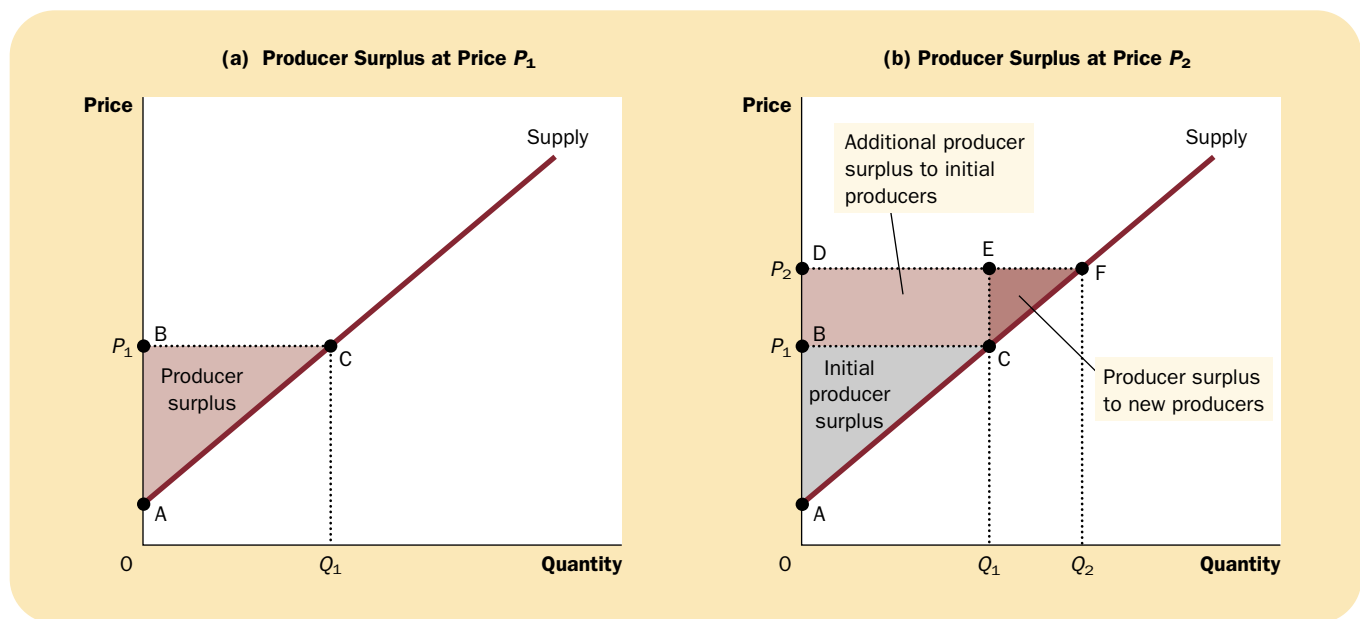
MEASURING PRODUCER SURPLUS WITH THE SUPPLY CURVE. In panel (a), the price of the good is \$600, and the producer surplus is \$100. In panel (b), the price of the good is \$800, and the producer surplus is \$500.

HOW A HIGHER PRICE RAISES PRODUCER SURPLUS

You will not be surprised to hear that sellers always want to receive a higher price for the goods they sell. But how much does sellers' well-being rise in response to a higher price? The concept of producer surplus offers a precise answer to this question.

Figure 7-6 shows a typical upward-sloping supply curve. Even though this supply curve differs in shape from the steplike supply curves in the previous figure, we measure producer surplus in the same way: Producer surplus is the area below the price and above the supply curve. In panel (a), the price is P_1 , and producer surplus is the area of triangle ABC.

Panel (b) shows what happens when the price rises from P_1 to P_2 . Producer surplus now equals area ADF. This increase in producer surplus has two parts. First, those sellers who were already selling Q_1 of the good at the lower price P_1 are better off because they now get more for what they sell. The increase in producer surplus for existing sellers equals the area of the rectangle BCED. Second, some new sellers enter the market because they are now willing to produce the good at the higher price, resulting in an increase in the quantity supplied from Q_1 to Q_2 . The producer surplus of these newcomers is the area of the triangle CEF.



HOW THE PRICE AFFECTS PRODUCER SURPLUS. In panel (a), the price is P_1 , the quantity demanded is Q_1 , and producer surplus equals the area of the triangle ABC. When the price rises from P_1 to P_2 , as in panel (b), the quantity supplied rises from Q_1 to Q_2 , and the producer surplus rises to the area of the triangle ADF. The increase in producer surplus (area BCFD) occurs in part because existing producers now receive more (area BCED) and in part because new producers enter the market at the higher price (area CEF).

Figure 7-6

As this analysis shows, we use producer surplus to measure the well-being of sellers in much the same way as we use consumer surplus to measure the well-being of buyers. Because these two measures of economic welfare are so similar, it is natural to use them together. And, indeed, that is exactly what we do in the next section.

QUICK QUIZ: Draw a supply curve for turkey. In your diagram, show a price of turkey and the producer surplus that results from that price. Explain in words what this producer surplus measures.

MARKET EFFICIENCY

Consumer surplus and producer surplus are the basic tools that economists use to study the welfare of buyers and sellers in a market. These tools can help us address a fundamental economic question: Is the allocation of resources determined by free markets in any way desirable?

THE BENEVOLENT SOCIAL PLANNER

To evaluate market outcomes, we introduce into our analysis a new, hypothetical character, called the benevolent social planner. The benevolent social planner is an all-knowing, all-powerful, well-intentioned dictator. The planner wants to maximize the economic well-being of everyone in society. What do you suppose this planner should do? Should he just leave buyers and sellers at the equilibrium that they reach naturally on their own? Or can he increase economic well-being by altering the market outcome in some way?

To answer this question, the planner must first decide how to measure the economic well-being of a society. One possible measure is the sum of consumer and producer surplus, which we call *total surplus*. Consumer surplus is the benefit that buyers receive from participating in a market, and producer surplus is the benefit that sellers receive. It is therefore natural to use total surplus as a measure of society's economic well-being.

To better understand this measure of economic well-being, recall how we measure consumer and producer surplus. We define consumer surplus as

$$\text{Consumer surplus} = \text{Value to buyers} - \text{Amount paid by buyers.}$$

Similarly, we define producer surplus as

$$\text{Producer surplus} = \text{Amount received by sellers} - \text{Cost to sellers.}$$

When we add consumer and producer surplus together, we obtain

$$\begin{aligned} \text{Total surplus} &= \text{Value to buyers} - \text{Amount paid by buyers} \\ &\quad + \text{Amount received by sellers} - \text{Cost to sellers.} \end{aligned}$$

The amount paid by buyers equals the amount received by sellers, so the middle two terms in this expression cancel each other. As a result, we can write total surplus as

$$\text{Total surplus} = \text{Value to buyers} - \text{Cost to sellers}.$$

Total surplus in a market is the total value to buyers of the goods, as measured by their willingness to pay, minus the total cost to sellers of providing those goods.

If an allocation of resources maximizes total surplus, we say that the allocation exhibits **efficiency**. If an allocation is not efficient, then some of the gains from trade among buyers and sellers are not being realized. For example, an allocation is inefficient if a good is not being produced by the sellers with lowest cost. In this case, moving production from a high-cost producer to a low-cost producer will lower the total cost to sellers and raise total surplus. Similarly, an allocation is inefficient if a good is not being consumed by the buyers who value it most highly. In this case, moving consumption of the good from a buyer with a low valuation to a buyer with a high valuation will raise total surplus.

In addition to efficiency, the social planner might also care about **equity**—the fairness of the distribution of well-being among the various buyers and sellers. In essence, the gains from trade in a market are like a pie to be distributed among the market participants. The question of efficiency is whether the pie is as big as possible. The question of equity is whether the pie is divided fairly. Evaluating the equity of a market outcome is more difficult than evaluating the efficiency. Whereas efficiency is an objective goal that can be judged on strictly positive grounds, equity involves normative judgments that go beyond economics and enter into the realm of political philosophy.

In this chapter we concentrate on efficiency as the social planner's goal. Keep in mind, however, that real policymakers often care about equity as well. That is, they care about both the size of the economic pie and how the pie gets sliced and distributed among members of society.

efficiency

the property of a resource allocation of maximizing the total surplus received by all members of society

equity

the fairness of the distribution of well-being among the members of society

EVALUATING THE MARKET EQUILIBRIUM

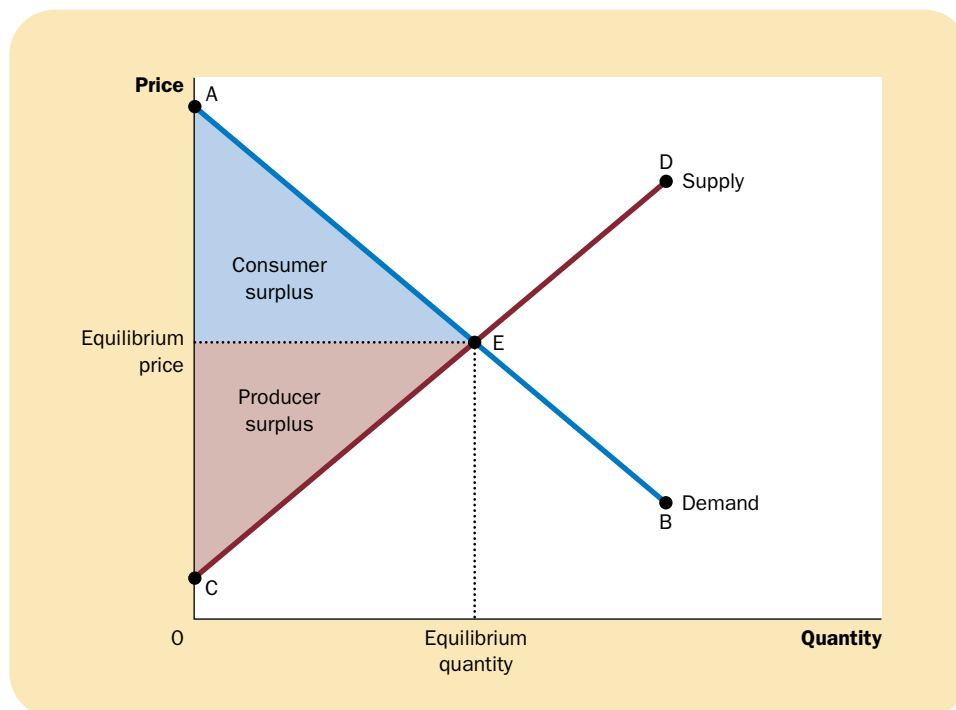
Figure 7-7 shows consumer and producer surplus when a market reaches the equilibrium of supply and demand. Recall that consumer surplus equals the area above the price and under the demand curve and producer surplus equals the area below the price and above the supply curve. Thus, the total area between the supply and demand curves up to the point of equilibrium represents the total surplus from this market.

Is this equilibrium allocation of resources efficient? Does it maximize total surplus? To answer these questions, keep in mind that when a market is in equilibrium, the price determines which buyers and sellers participate in the market. Those buyers who value the good more than the price (represented by the segment AE on the demand curve) choose to buy the good; those buyers who value it less than the price (represented by the segment EB) do not. Similarly, those sellers whose costs are less than the price (represented by the segment CE on the supply curve) choose to produce and sell the good; those sellers whose costs are greater than the price (represented by the segment ED) do not.

These observations lead to two insights about market outcomes:

Figure 7-7

CONSUMER AND PRODUCER SURPLUS IN THE MARKET EQUILIBRIUM. Total surplus—the sum of consumer and producer surplus—is the area between the supply and demand curves up to the equilibrium quantity.



1. Free markets allocate the supply of goods to the buyers who value them most highly, as measured by their willingness to pay.
2. Free markets allocate the demand for goods to the sellers who can produce them at least cost.

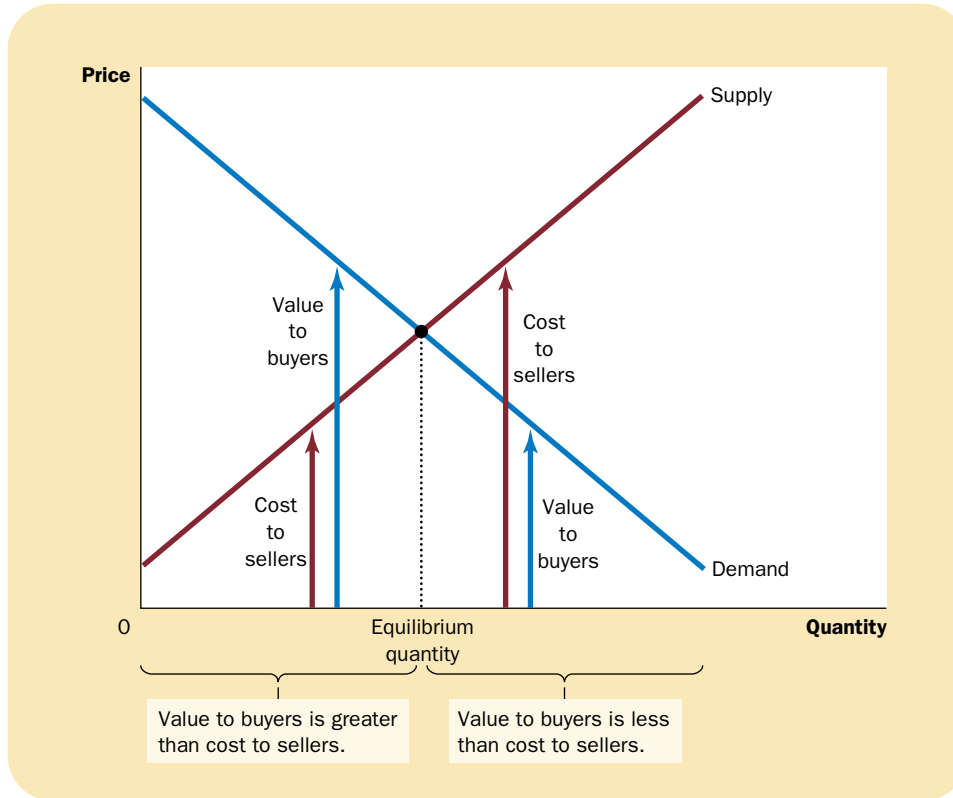
Thus, given the quantity produced and sold in a market equilibrium, the social planner cannot increase economic well-being by changing the allocation of consumption among buyers or the allocation of production among sellers.

But can the social planner raise total economic well-being by increasing or decreasing the quantity of the good? The answer is no, as stated in this third insight about market outcomes:

3. Free markets produce the quantity of goods that maximizes the sum of consumer and producer surplus.

To see why this is true, consider Figure 7-8. Recall that the demand curve reflects the value to buyers and that the supply curve reflects the cost to sellers. At quantities below the equilibrium level, the value to buyers exceeds the cost to sellers. In this region, increasing the quantity raises total surplus, and it continues to do so until the quantity reaches the equilibrium level. Beyond the equilibrium quantity, however, the value to buyers is less than the cost to sellers. Producing more than the equilibrium quantity would, therefore, lower total surplus.

These three insights about market outcomes tell us that the equilibrium of supply and demand maximizes the sum of consumer and producer surplus. In other words, the equilibrium outcome is an efficient allocation of resources. The job of the benevolent social planner is, therefore, very easy: He can leave the market

**Figure 7-8**

THE EFFICIENCY OF THE EQUILIBRIUM QUANTITY. At quantities less than the equilibrium quantity, the value to buyers exceeds the cost to sellers. At quantities greater than the equilibrium quantity, the cost to sellers exceeds the value to buyers. Therefore, the market equilibrium maximizes the sum of producer and consumer surplus.

outcome just as he finds it. This policy of leaving well enough alone goes by the French expression *laissez-faire*, which literally translated means “allow them to do.”

We can now better appreciate Adam Smith’s invisible hand of the marketplace, which we first discussed in Chapter 1. The benevolent social planner doesn’t need to alter the market outcome because the invisible hand has already guided buyers and sellers to an allocation of the economy’s resources that maximizes total surplus. This conclusion explains why economists often advocate free markets as the best way to organize economic activity.

QUICK QUIZ: Draw the supply and demand for turkey. In the equilibrium, show producer and consumer surplus. Explain why producing more turkey would lower total surplus.

CONCLUSION: MARKET EFFICIENCY AND MARKET FAILURE

This chapter introduced the basic tools of welfare economics—consumer and producer surplus—and used them to evaluate the efficiency of free markets. We showed that the forces of supply and demand allocate resources efficiently. That is,

IN THE NEWS

Ticket Scalping



IF AN ECONOMY IS TO ALLOCATE ITS SCARCE resources efficiently, goods must get to those consumers who value them most highly. Ticket scalping is one example of how markets reach efficient outcomes. Scalpers buy tickets to plays, concerts, and sports events and then sell the tickets at a price above their original cost. By charging the highest price the market will bear, scalpers help ensure that consumers with the greatest willingness to pay for the tickets actually do get them. In some places, however, there is debate over whether this market activity should be legal.

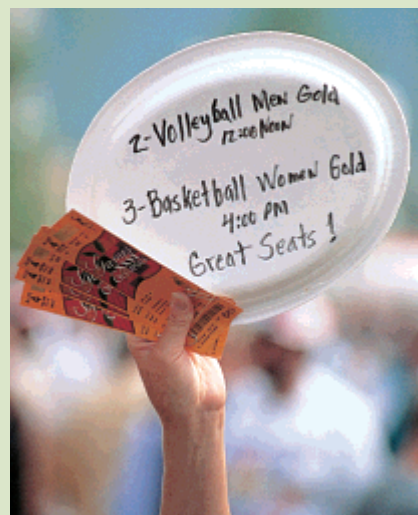
Tickets? Supply Meets Demand on Sidewalk

BY JOHN TIERNEY

Ticket scalping has been very good to Kevin Thomas, and he makes no apologies. He sees himself as a classic American entrepreneur: a high school dropout from the Bronx who taught himself a trade, works seven nights a week, earns \$40,000 a year, and at age twenty-six has \$75,000 in savings, all by providing a public service outside New York's theaters and sports arenas.

He has just one complaint. "I've been busted about 30 times in the last year," he said one recent evening, just after making \$280 at a Knicks game. "You learn to deal with it—I give the cops a fake name, and I pay the fines when I have to, but I don't think it's fair. I look at scalping like working as a stockbroker, buying low and selling high. If people are willing to pay me the money, what kind of problem is that?"

It is a significant problem to public officials in New York and New Jersey,



THE INVISIBLE HAND AT WORK

who are cracking down on street scalpers like Mr. Thomas and on licensed ticket brokers. Undercover officers are enforcing new restrictions on reselling tickets at marked-up prices, and the attorneys general of the two states are pressing well-publicized

even though each buyer and seller in a market is concerned only about his or her own welfare, they are together led by an invisible hand to an equilibrium that maximizes the total benefits to buyers and sellers.

A word of warning is in order. To conclude that markets are efficient, we made several assumptions about how markets work. When these assumptions do not hold, our conclusion that the market equilibrium is efficient may no longer be true. As we close this chapter, let's consider briefly two of the most important of these assumptions.

First, our analysis assumed that markets are perfectly competitive. In the world, however, competition is sometimes far from perfect. In some markets, a single buyer or seller (or a small group of them) may be able to control market prices. This ability to influence prices is called *market power*. Market power can cause markets to be inefficient because it keeps the price and quantity away from the equilibrium of supply and demand.

Second, our analysis assumed that the outcome in a market matters only to the buyers and sellers in that market. Yet, in the world, the decisions of buyers and

cases against more than a dozen ticket brokers.

But economists tend to see scalping from Mr. Thomas's perspective. To them, the governments' crusade makes about as much sense as the old campaigns by Communist authorities against "profiteering." Economists argue that the restrictions inconvenience the public, reduce the audience for cultural and sports events, waste the police's time, deprive New York City of tens of millions of dollars of tax revenue, and actually drive up the cost of many tickets.

"It is always good politics to pose as defender of the poor by declaring high prices illegal," says William J. Baumol, the director of the C. V. Starr Center for Applied Economics at New York University. "I expect politicians to try to solve the AIDS crisis by declaring AIDS illegal as well. That would be harmless, because nothing would happen, but when you outlaw high prices you create real problems."

Dr. Baumol was one of the economists who came up with the idea of sell-

ing same-day Broadway tickets for half price at the TKTS booth in Times Square, which theater owners thought dangerously radical when the booth opened in 1973. But the owners have profited by finding a new clientele for tickets that would have gone unsold, an illustration of the free-market tenet that both buyers and sellers ultimately benefit when price is adjusted to meet demand.

Economists see another illustration of that lesson at the Museum of Modern Art, where people wait in line for up to two hours to buy tickets for the Matisse exhibit. But there is an alternative on the sidewalk: Scalpers who evade the police have been selling the \$12.50 tickets to the show at prices ranging from \$20 to \$50.

"You don't have to put a very high value on your time to pay \$10 or \$15 to avoid standing in line for two hours for a Matisse ticket," said Richard H. Thaler, an economist at Cornell University. "Some people think it's fairer to make everyone stand in line, but that forces everyone to engage in a totally unpro-

ductive activity, and it discriminates in favor of people who have the most free time. Scalping gives other people a chance, too. I can see no justification for outlawing it." . . .

Politicians commonly argue that without anti-scalping laws, tickets would become unaffordable to most people, but California has no laws against scalping, and ticket prices there are not notoriously high. And as much as scalpers would like to inflate prices, only a limited number of people are willing to pay \$100 for a ticket. . . .

Legalizing scalping, however, would not necessarily be good news for everyone. Mr. Thomas, for instance, fears that the extra competition might put him out of business. But after 16 years—he started at age ten outside of Yankee Stadium—he is thinking it might be time for a change anyway.

SOURCE: *The New York Times*, December 26, 1992, p. A1.

sellers sometimes affect people who are not participants in the market at all. Pollution is the classic example of a market outcome that affects people not in the market. Such side effects, called *externalities*, cause welfare in a market to depend on more than just the value to the buyers and the cost to the sellers. Because buyers and sellers do not take these side effects into account when deciding how much to consume and produce, the equilibrium in a market can be inefficient from the standpoint of society as a whole.

Market power and externalities are examples of a general phenomenon called *market failure*—the inability of some unregulated markets to allocate resources efficiently. When markets fail, public policy can potentially remedy the problem and increase economic efficiency. Microeconomists devote much effort to studying when market failure is likely and what sorts of policies are best at correcting market failures. As you continue your study of economics, you will see that the tools of welfare economics developed here are readily adapted to that endeavor.

Despite the possibility of market failure, the invisible hand of the marketplace is extraordinarily important. In many markets, the assumptions we made in this

chapter work well, and the conclusion of market efficiency applies directly. Moreover, our analysis of welfare economics and market efficiency can be used to shed light on the effects of various government policies. In the next two chapters we apply the tools we have just developed to study two important policy issues—the welfare effects of taxation and of international trade.

Summary

- ◆ Consumer surplus equals buyers' willingness to pay for a good minus the amount they actually pay for it, and it measures the benefit buyers get from participating in a market. Consumer surplus can be computed by finding the area below the demand curve and above the price.
- ◆ Producer surplus equals the amount sellers receive for their goods minus their costs of production, and it measures the benefit sellers get from participating in a market. Producer surplus can be computed by finding the area below the price and above the supply curve.
- ◆ An allocation of resources that maximizes the sum of consumer and producer surplus is said to be efficient.
- ◆ Policymakers are often concerned with the efficiency, as well as the equity, of economic outcomes.
- ◆ The equilibrium of supply and demand maximizes the sum of consumer and producer surplus. That is, the invisible hand of the marketplace leads buyers and sellers to allocate resources efficiently.
- ◆ Markets do not allocate resources efficiently in the presence of market failures such as market power or externalities.

Key Concepts

welfare economics, p. 142
willingness to pay, p. 142
consumer surplus, p. 143

cost, p. 148
producer surplus, p. 148

efficiency, p. 153
equity, p. 153

Questions for Review

1. Explain how buyers' willingness to pay, consumer surplus, and the demand curve are related.
2. Explain how sellers' costs, producer surplus, and the supply curve are related.
3. In a supply-and-demand diagram, show producer and consumer surplus in the market equilibrium.
4. What is efficiency? Is it the only goal of economic policymakers?
5. What does the invisible hand do?
6. Name two types of market failure. Explain why each may cause market outcomes to be inefficient.

Problems and Applications

1. An early freeze in California sours the lemon crop. What happens to consumer surplus in the market for lemons? What happens to consumer surplus in the market for lemonade? Illustrate your answers with diagrams.
2. Suppose the demand for French bread rises. What happens to producer surplus in the market for French bread? What happens to producer surplus in the market for flour? Illustrate your answer with diagrams.

3. It is a hot day, and Bert is very thirsty. Here is the value he places on a bottle of water:

Value of first bottle	\$7
Value of second bottle	5
Value of third bottle	3
Value of fourth bottle	1

- From this information, derive Bert's demand schedule. Graph his demand curve for bottled water.
 - If the price of a bottle of water is \$4, how many bottles does Bert buy? How much consumer surplus does Bert get from his purchases? Show Bert's consumer surplus in your graph.
 - If the price falls to \$2, how does quantity demanded change? How does Bert's consumer surplus change? Show these changes in your graph.
4. Ernie owns a water pump. Because pumping large amounts of water is harder than pumping small amounts, the cost of producing a bottle of water rises as he pumps more. Here is the cost he incurs to produce each bottle of water:

Cost of first bottle	\$1
Cost of second bottle	3
Cost of third bottle	5
Cost of fourth bottle	7

- From this information, derive Ernie's supply schedule. Graph his supply curve for bottled water.
 - If the price of a bottle of water is \$4, how many bottles does Ernie produce and sell? How much producer surplus does Ernie get from these sales? Show Ernie's producer surplus in your graph.
 - If the price rises to \$6, how does quantity supplied change? How does Ernie's producer surplus change? Show these changes in your graph.
5. Consider a market in which Bert from Problem 3 is the buyer and Ernie from Problem 4 is the seller.
- Use Ernie's supply schedule and Bert's demand schedule to find the quantity supplied and quantity demanded at prices of \$2, \$4, and \$6. Which of these prices brings supply and demand into equilibrium?
 - What are consumer surplus, producer surplus, and total surplus in this equilibrium?
 - If Ernie produced and Bert consumed one less bottle of water, what would happen to total surplus?

- If Ernie produced and Bert consumed one additional bottle of water, what would happen to total surplus?

6. The cost of producing stereo systems has fallen over the past several decades. Let's consider some implications of this fact.
- Use a supply-and-demand diagram to show the effect of falling production costs on the price and quantity of stereos sold.
 - In your diagram, show what happens to consumer surplus and producer surplus.
 - Suppose the supply of stereos is very elastic. Who benefits most from falling production costs—consumers or producers of stereos?
7. There are four consumers willing to pay the following amounts for haircuts:

Jerry: \$7 Oprah: \$2 Sally Jessy: \$8 Montel: \$5

There are four haircutting businesses with the following costs:

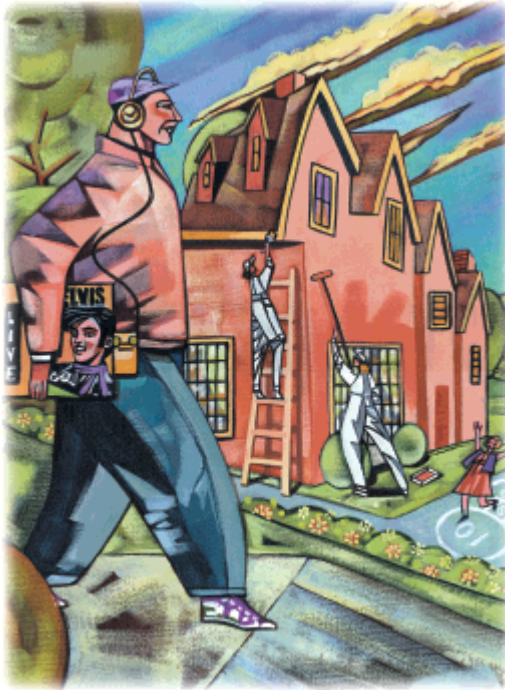
Firm A: \$3 Firm B: \$6 Firm C: \$4 Firm D: \$2

Each firm has the capacity to produce only one haircut. For efficiency, how many haircuts should be given? Which businesses should cut hair, and which consumers should have their hair cut? How large is the maximum possible total surplus?

8. Suppose a technological advance reduces the cost of making computers.
- Use a supply-and-demand diagram to show what happens to price, quantity, consumer surplus, and producer surplus in the market for computers.
 - Computers and adding machines are substitutes. Use a supply-and-demand diagram to show what happens to price, quantity, consumer surplus, and producer surplus in the market for adding machines. Should adding machine producers be happy or sad about the technological advance in computers?
 - Computers and software are complements. Use a supply-and-demand diagram to show what happens to price, quantity, consumer surplus, and producer surplus in the market for software. Should software producers be happy or sad about the technological advance in computers?
 - Does this analysis help explain why Bill Gates, a software producer, is one of the world's richest men?

9. Consider how health insurance affects the quantity of health care services performed. Suppose that the typical medical procedure has a cost of \$100, yet a person with health insurance pays only \$20 out-of-pocket when she chooses to have an additional procedure performed. Her insurance company pays the remaining \$80. (The insurance company will recoup the \$80 through higher premiums for everybody, but the share paid by this individual is small.)
 - a. Draw the demand curve in the market for medical care. (In your diagram, the horizontal axis should represent the number of medical procedures.) Show the quantity of procedures demanded if each procedure has a price of \$100.
 - b. On your diagram, show the quantity of procedures demanded if consumers pay only \$20 per procedure. If the cost of each procedure to society is truly \$100, and if individuals have health insurance as just described, will the number of procedures performed maximize total surplus? Explain.
 - c. Economists often blame the health insurance system for excessive use of medical care. Given your analysis, why might the use of care be viewed as “excessive”?
 - d. What sort of policies might prevent this excessive use?
10. Many parts of California experienced a severe drought in the late 1980s and early 1990s.
 - a. Use a diagram of the water market to show the effects of the drought on the equilibrium price and quantity of water.
 - b. Many communities did not allow the price of water to change, however. What is the effect of this policy on the water market? Show on your diagram any surplus or shortage that arises.
 - c. A 1991 op-ed piece in *The Wall Street Journal* stated that “all Los Angeles residents are required to cut their water usage by 10 percent as of March 1 and another 5 percent starting May 1, based on their 1986 consumption levels.” The author criticized this policy on both efficiency and equity grounds, saying “not only does such a policy reward families who ‘wasted’ more water back in 1986, it does little to encourage consumers who could make more drastic reductions, [and] . . . punishes consumers who cannot so readily reduce their water use.” In what way is the Los Angeles system for allocating water inefficient? In what way does the system seem unfair?
 - d. Suppose instead that Los Angeles allowed the price of water to increase until the quantity demanded equaled the quantity supplied. Would the resulting allocation of water be more efficient? In your view, would it be more or less fair than the proportionate reductions in water use mentioned in the newspaper article? What could be done to make the market solution more fair?

8



APPLICATION: THE COSTS OF TAXATION

Taxes are often a source of heated political debate. In 1776 the anger of the American colonies over British taxes sparked the American Revolution. More than two centuries later Ronald Reagan was elected president on a platform of large cuts in personal income taxes, and during his eight years in the White House the top tax rate on income fell from 70 percent to 28 percent. In 1992 Bill Clinton was elected in part because incumbent George Bush had broken his 1988 campaign promise, "Read my lips: no new taxes."

We began our study of taxes in Chapter 6. There we saw how a tax on a good affects its price and the quantity sold and how the forces of supply and demand divide the burden of a tax between buyers and sellers. In this chapter we extend this analysis and look at how taxes affect welfare, the economic well-being of participants in a market.

**IN THIS CHAPTER
YOU WILL . . .**

***Examine how taxes
reduce consumer
and producer
surplus***

***Learn the meaning
and causes of the
deadweight loss of
a tax***

***Consider why some
taxes have larger
deadweight losses
than others***

***Examine how tax
revenue and
deadweight loss
vary with the size
of a tax***



"You know, the idea of taxation *with* representation doesn't appeal to me very much, either."

The effects of taxes on welfare might at first seem obvious. The government enacts taxes to raise revenue, and that revenue must come out of someone's pocket. As we saw in Chapter 6, both buyers and sellers are worse off when a good is taxed: A tax raises the price buyers pay and lowers the price sellers receive. Yet to understand fully how taxes affect economic well-being, we must compare the reduced welfare of buyers and sellers to the amount of revenue the government raises. The tools of consumer and producer surplus allow us to make this comparison. The analysis will show that the costs of taxes to buyers and sellers exceeds the revenue raised by the government.

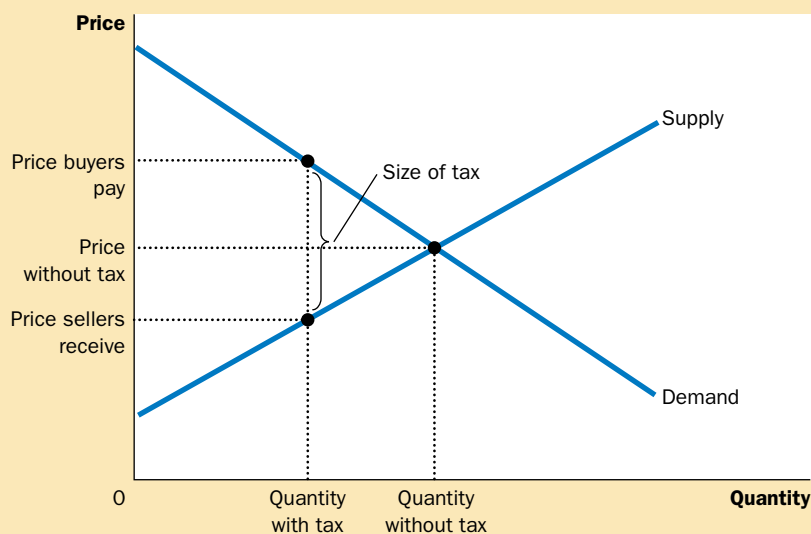
THE DEADWEIGHT LOSS OF TAXATION

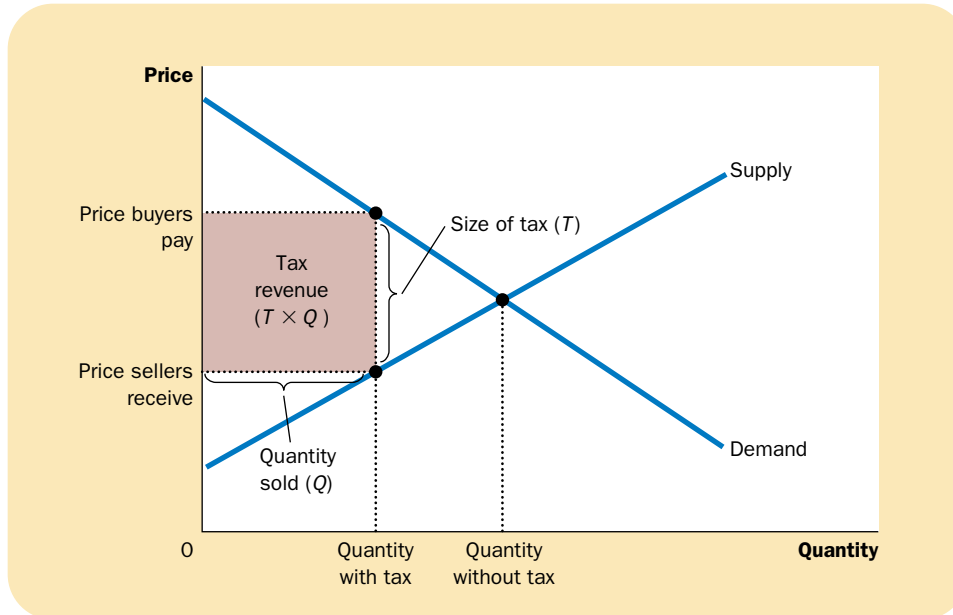
We begin by recalling one of the surprising lessons from Chapter 6: It does not matter whether a tax on a good is levied on buyers or sellers of the good. When a tax is levied on buyers, the demand curve shifts downward by the size of the tax; when it is levied on sellers, the supply curve shifts upward by that amount. In either case, when the tax is enacted, the price paid by buyers rises, and the price received by sellers falls. In the end, buyers and sellers share the burden of the tax, regardless of how it is levied.

Figure 8-1 shows these effects. To simplify our discussion, this figure does not show a shift in either the supply or demand curve, although one curve must shift. Which curve shifts depends on whether the tax is levied on sellers (the supply curve shifts) or buyers (the demand curve shifts). In this chapter, we can simplify the graphs by not bothering to show the shift. The key result for our purposes here

Figure 8-1

THE EFFECTS OF A TAX. A tax on a good places a wedge between the price that buyers pay and the price that sellers receive. The quantity of the good sold falls.



**Figure 8-2**

TAX REVENUE. The tax revenue that the government collects equals $T \times Q$, the size of the tax T times the quantity sold Q . Thus, tax revenue equals the area of the rectangle between the supply and demand curves.

is that the tax places a wedge between the price buyers pay and the price sellers receive. Because of this tax wedge, the quantity sold falls below the level that would be sold without a tax. In other words, a tax on a good causes the size of the market for the good to shrink. These results should be familiar from Chapter 6.

HOW A TAX AFFECTS MARKET PARTICIPANTS

Now let's use the tools of welfare economics to measure the gains and losses from a tax on a good. To do this, we must take into account how the tax affects buyers, sellers, and the government. The benefit received by buyers in a market is measured by consumer surplus—the amount buyers are willing to pay for the good minus the amount they actually pay for it. The benefit received by sellers in a market is measured by producer surplus—the amount sellers receive for the good minus their costs. These are precisely the measures of economic welfare we used in Chapter 7.

What about the third interested party, the government? If T is the size of the tax and Q is the quantity of the good sold, then the government gets total tax revenue of $T \times Q$. It can use this tax revenue to provide services, such as roads, police, and public education, or to help the needy. Therefore, to analyze how taxes affect economic well-being, we use tax revenue to measure the government's benefit from the tax. Keep in mind, however, that this benefit actually accrues not to government but to those on whom the revenue is spent.

Figure 8-2 shows that the government's tax revenue is represented by the rectangle between the supply and demand curves. The height of this rectangle is the size of the tax, T , and the width of the rectangle is the quantity of the good sold, Q . Because a rectangle's area is its height times its width, this rectangle's area is $T \times Q$, which equals the tax revenue.

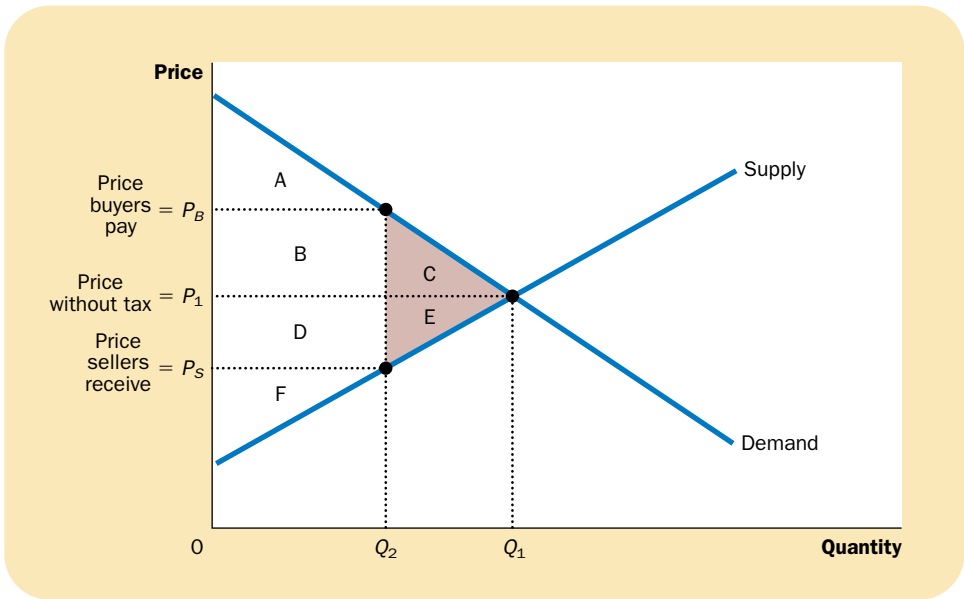
Welfare without a Tax To see how a tax affects welfare, we begin by considering welfare before the government has imposed a tax. Figure 8-3 shows the supply-and-demand diagram and marks the key areas with the letters A through F.

Without a tax, the price and quantity are found at the intersection of the supply and demand curves. The price is P_1 , and the quantity sold is Q_1 . Because the demand curve reflects buyers' willingness to pay, consumer surplus is the area between the demand curve and the price, $A + B + C$. Similarly, because the supply curve reflects sellers' costs, producer surplus is the area between the supply curve and the price, $D + E + F$. In this case, because there is no tax, tax revenue equals zero.

Total surplus, the sum of consumer and producer surplus, equals the area $A + B + C + D + E + F$. In other words, as we saw in Chapter 7, total surplus is the area between the supply and demand curves up to the equilibrium quantity. The first column of Table 8-1 summarizes these conclusions.

Figure 8-3

HOW A TAX AFFECTS WELFARE. A tax on a good reduces consumer surplus (by the area $B + C$) and producer surplus (by the area $D + E$). Because the fall in producer and consumer surplus exceeds tax revenue (area $B + D$), the tax is said to impose a deadweight loss (area $C + E$).



	WITHOUT TAX	WITH TAX	CHANGE
Consumer Surplus	$A + B + C$	A	$-(B + C)$
Producer Surplus	$D + E + F$	F	$-(D + E)$
Tax Revenue	None	$B + D$	$+(B + D)$
Total Surplus	$A + B + C + D + E + F$	$A + B + D + F$	$-(C + E)$

The area $C + E$ shows the fall in total surplus and is the deadweight loss of the tax.

Table 8-1

CHANGES IN WELFARE FROM A TAX. This table refers to the areas marked in Figure 8-3 to show how a tax affects the welfare of buyers and sellers in a market.

Welfare with a Tax Now consider welfare after the tax is enacted. The price paid by buyers rises from P_1 to P_B , so consumer surplus now equals only area A (the area below the demand curve and above the buyer's price). The price received by sellers falls from P_1 to P_S , so producer surplus now equals only area F (the area above the supply curve and below the seller's price). The quantity sold falls from Q_1 to Q_2 , and the government collects tax revenue equal to the area B + D.

To compute total surplus with the tax, we add consumer surplus, producer surplus, and tax revenue. Thus, we find that total surplus is area A + B + D + F. The second column of Table 8-1 provides a summary.

Changes in Welfare We can now see the effects of the tax by comparing welfare before and after the tax is enacted. The third column in Table 8-1 shows the changes. The tax causes consumer surplus to fall by the area B + C and producer surplus to fall by the area D + E. Tax revenue rises by the area B + D. Not surprisingly, the tax makes buyers and sellers worse off and the government better off.

The change in total welfare includes the change in consumer surplus (which is negative), the change in producer surplus (which is also negative), and the change in tax revenue (which is positive). When we add these three pieces together, we find that total surplus in the market falls by the area C + E. Thus, the losses to buyers and sellers from a tax exceed the revenue raised by the government. The fall in total surplus that results when a tax (or some other policy) distorts a market outcome is called the **deadweight loss**. The area C + E measures the size of the deadweight loss.

To understand why taxes impose deadweight losses, recall one of the *Ten Principles of Economics* in Chapter 1: People respond to incentives. In Chapter 7 we saw that markets normally allocate scarce resources efficiently. That is, the equilibrium of supply and demand maximizes the total surplus of buyers and sellers in a market. When a tax raises the price to buyers and lowers the price to sellers, however, it gives buyers an incentive to consume less and sellers an incentive to produce less than they otherwise would. As buyers and sellers respond to these incentives, the size of the market shrinks below its optimum. Thus, because taxes distort incentives, they cause markets to allocate resources inefficiently.

deadweight loss

the fall in total surplus that results from a market distortion, such as a tax



DEADWEIGHT LOSSES AND THE GAINS FROM TRADE

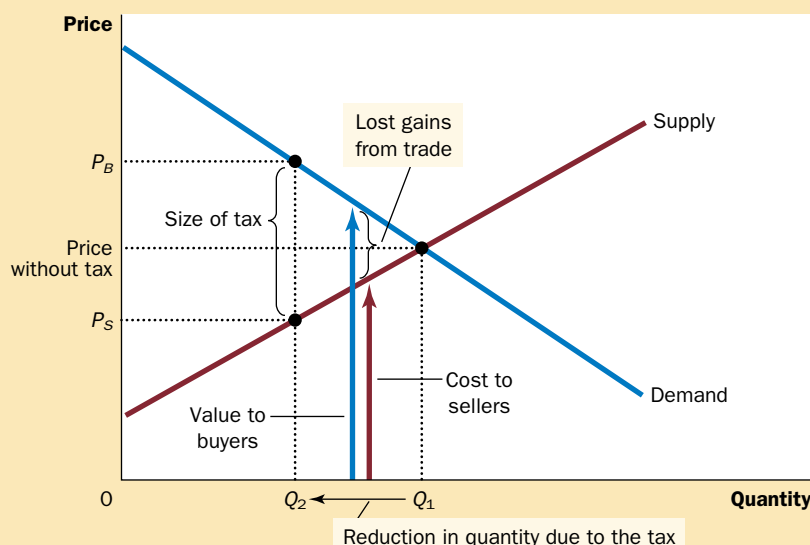
To gain some intuition for why taxes result in deadweight losses, consider an example. Imagine that Joe cleans Jane's house each week for \$100. The opportunity cost of Joe's time is \$80, and the value of a clean house to Jane is \$120. Thus, Joe and Jane each receive a \$20 benefit from their deal. The total surplus of \$40 measures the gains from trade in this particular transaction.

Now suppose that the government levies a \$50 tax on the providers of cleaning services. There is now no price that Jane can pay Joe that will leave both of them better off after paying the tax. The most Jane would be willing to pay is \$120, but then Joe would be left with only \$70 after paying the tax, which is less than his \$80 opportunity cost. Conversely, for Joe to receive his opportunity cost of \$80, Jane would need to pay \$130, which is above the \$120 value she places on a clean house. As a result, Jane and Joe cancel their arrangement. Joe goes without the income, and Jane lives in a dirtier house.

The tax has made Joe and Jane worse off by a total of \$40, because they have lost this amount of surplus. At the same time, the government collects no revenue from Joe and Jane because they decide to cancel their arrangement. The \$40 is pure

Figure 8-4

THE DEADWEIGHT LOSS. When the government imposes a tax on a good, the quantity sold falls from Q_1 to Q_2 . As a result, some of the potential gains from trade among buyers and sellers do not get realized. These lost gains from trade create the deadweight loss.



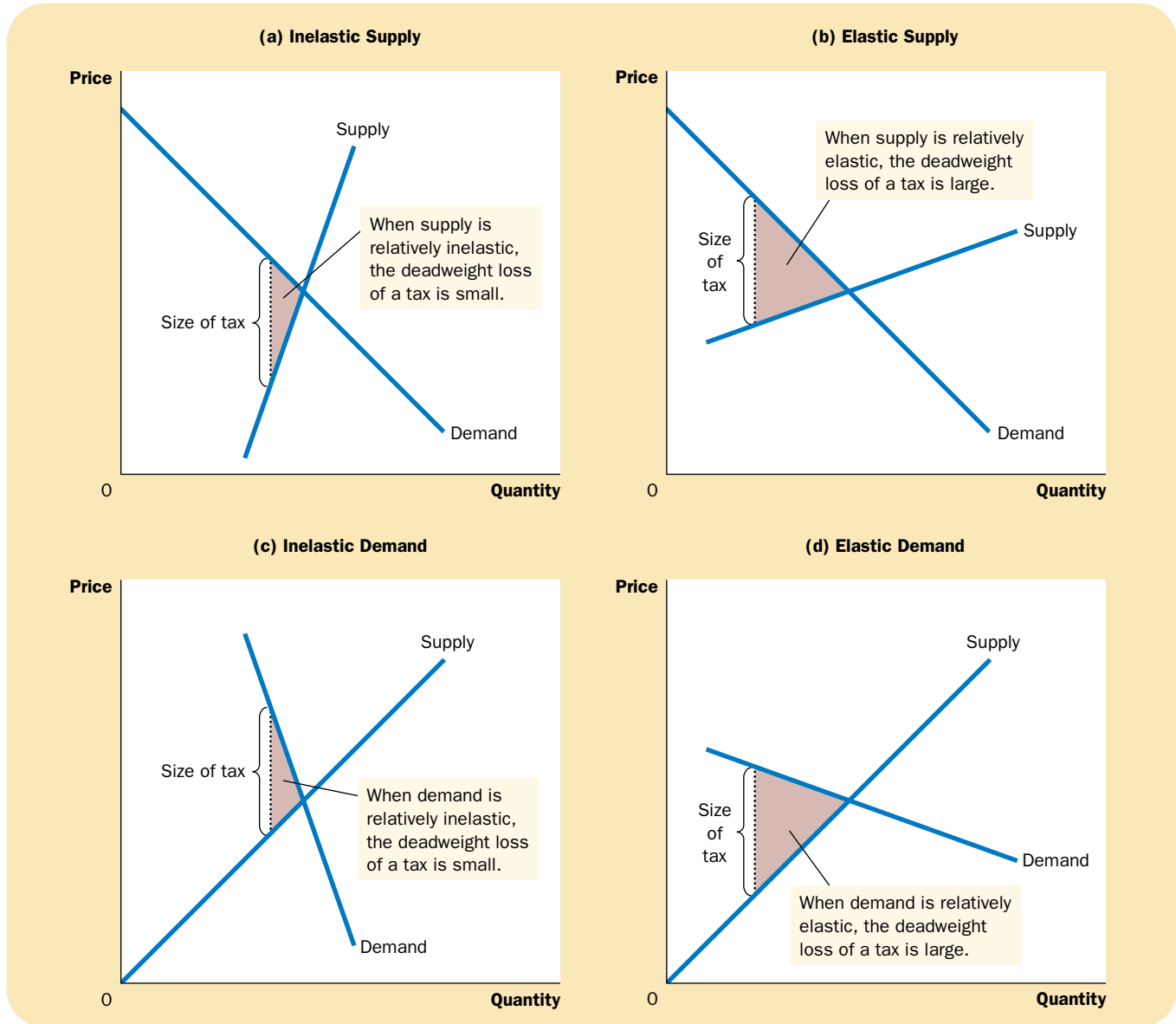
deadweight loss: It is a loss to buyers and sellers in a market not offset by an increase in government revenue. From this example, we can see the ultimate source of deadweight losses: *Taxes cause deadweight losses because they prevent buyers and sellers from realizing some of the gains from trade.*

The area of the triangle between the supply and demand curves (area C + E in Figure 8-3) measures these losses. This loss can be seen most easily in Figure 8-4 by recalling that the demand curve reflects the value of the good to consumers and that the supply curve reflects the costs of producers. When the tax raises the price to buyers to P_B and lowers the price to sellers to P_S , the marginal buyers and sellers leave the market, so the quantity sold falls from Q_1 to Q_2 . Yet, as the figure shows, the value of the good to these buyers still exceeds the cost to these sellers. As in our example with Joe and Jane, the gains from trade—the difference between buyers' value and sellers' cost—is less than the tax. Thus, these trades do not get made once the tax is imposed. The deadweight loss is the surplus lost because the tax discourages these mutually advantageous trades.

QUICK QUIZ: Draw the supply and demand curve for cookies. If the government imposes a tax on cookies, show what happens to the quantity sold, the price paid by buyers, and the price paid by sellers. In your diagram, show the deadweight loss from the tax. Explain the meaning of the deadweight loss.

THE DETERMINANTS OF THE DEADWEIGHT LOSS

What determines whether the deadweight loss from a tax is large or small? The answer is the price elasticities of supply and demand, which measure how much the quantity supplied and quantity demanded respond to changes in the price.



TAX DISTORTIONS AND ELASTICITIES. In panels (a) and (b), the demand curve and the size of the tax are the same, but the price elasticity of supply is different. Notice that the more elastic the supply curve, the larger the deadweight loss of the tax. In panels (c) and (d), the supply curve and the size of the tax are the same, but the price elasticity of demand is different. Notice that the more elastic the demand curve, the larger the deadweight loss of the tax.

Figure 8-5

Let's consider first how the elasticity of supply affects the size of the deadweight loss. In the top two panels of Figure 8-5, the demand curve and the size of the tax are the same. The only difference in these figures is the elasticity of the supply curve. In panel (a), the supply curve is relatively inelastic: Quantity supplied responds only slightly to changes in the price. In panel (b), the supply curve is

relatively elastic: Quantity supplied responds substantially to changes in the price. Notice that the deadweight loss, the area of the triangle between the supply and demand curves, is larger when the supply curve is more elastic.

Similarly, the bottom two panels of Figure 8-5 show how the elasticity of demand affects the size of the deadweight loss. Here the supply curve and the size of the tax are held constant. In panel (c) the demand curve is relatively inelastic, and the deadweight loss is small. In panel (d) the demand curve is more elastic, and the deadweight loss from the tax is larger.

The lesson from this figure is easy to explain. A tax has a deadweight loss because it induces buyers and sellers to change their behavior. The tax raises the price paid by buyers, so they consume less. At the same time, the tax lowers the price received by sellers, so they produce less. Because of these changes in behavior, the size of the market shrinks below the optimum. The elasticities of supply and demand measure how much sellers and buyers respond to the changes in the price and, therefore, determine how much the tax distorts the market outcome. Hence, *the greater the elasticities of supply and demand, the greater the deadweight loss of a tax.*

CASE STUDY THE DEADWEIGHT LOSS DEBATE

Supply, demand, elasticity, deadweight loss—all this economic theory is enough to make your head spin. But believe it or not, these ideas go to the heart of a profound political question: How big should the government be? The reason the debate hinges on these concepts is that the larger the deadweight loss of taxation, the larger the cost of any government program. If taxation entails very large deadweight losses, then these losses are a strong argument for a leaner government that does less and taxes less. By contrast, if taxes impose only small deadweight losses, then government programs are less costly than they otherwise might be.

So how big are the deadweight losses of taxation? This is a question about which economists disagree. To see the nature of this disagreement, consider the most important tax in the U.S. economy—the tax on labor. The Social Security tax, the Medicare tax, and, to a large extent, the federal income tax are labor taxes. Many state governments also tax labor earnings. A labor tax places a wedge between the wage that firms pay and the wage that workers receive. If we add all forms of labor taxes together, the *marginal tax rate* on labor income—the tax on the last dollar of earnings—is almost 50 percent for many workers.

Although the size of the labor tax is easy to determine, the deadweight loss of this tax is less straightforward. Economists disagree about whether this 50 percent labor tax has a small or a large deadweight loss. This disagreement arises because they hold different views about the elasticity of labor supply.

Economists who argue that labor taxes are not very distorting believe that labor supply is fairly inelastic. Most people, they claim, would work full-time regardless of the wage. If so, the labor supply curve is almost vertical, and a tax on labor has a small deadweight loss.

Economists who argue that labor taxes are highly distorting believe that labor supply is more elastic. They admit that some groups of workers may supply their labor inelastically but claim that many other groups respond more to incentives. Here are some examples:

- ◆ Many workers can adjust the number of hours they work—for instance, by working overtime. The higher the wage, the more hours they choose to work.



"LET ME TELL YOU WHAT I THINK ABOUT THE ELASTICITY OF LABOR SUPPLY."

- ◆ Some families have second earners—often married women with children—with some discretion over whether to do unpaid work at home or paid work in the marketplace. When deciding whether to take a job, these second earners compare the benefits of being at home (including savings on the cost of child care) with the wages they could earn.
- ◆ Many of the elderly can choose when to retire, and their decisions are partly based on the wage. Once they are retired, the wage determines their incentive to work part-time.
- ◆ Some people consider engaging in illegal economic activity, such as the drug trade, or working at jobs that pay “under the table” to evade taxes. Economists call this the *underground economy*. In deciding whether to work in the underground economy or at a legitimate job, these potential criminals compare what they can earn by breaking the law with the wage they can earn legally.

In each of these cases, the quantity of labor supplied responds to the wage (the price of labor). Thus, the decisions of these workers are distorted when their labor earnings are taxed. Labor taxes encourage workers to work fewer hours, second earners to stay at home, the elderly to retire early, and the unscrupulous to enter the underground economy.

These two views of labor taxation persist to this day. Indeed, whenever you see two political candidates debating whether the government should provide more services or reduce the tax burden, keep in mind that part of the disagreement may rest on different views about the elasticity of labor supply and the deadweight loss of taxation.

QUICK QUIZ: The demand for beer is more elastic than the demand for milk. Would a tax on beer or a tax on milk have larger deadweight loss? Why?

FYI

Henry George
and the
Land Tax

mayor of New York City (although he finished well ahead of Republican candidate Theodore Roosevelt).

George's proposal to tax land was motivated largely by a concern over the distribution of economic well-being. He deplored the "shocking contrast between monstrous wealth and debasing want" and thought landowners benefited more than they should from the rapid growth in the overall economy.

George's arguments for the land tax can be understood using the tools of modern economics. Consider first supply and demand in the market for renting land. As immigration causes the population to rise and technological progress causes incomes to grow, the demand for land rises over time. Yet because the amount of land is fixed, the supply is perfectly inelastic. Rapid increases in demand together with inelastic supply lead to large increases in the equilibrium rents on land, so that economic growth makes rich landowners even richer.

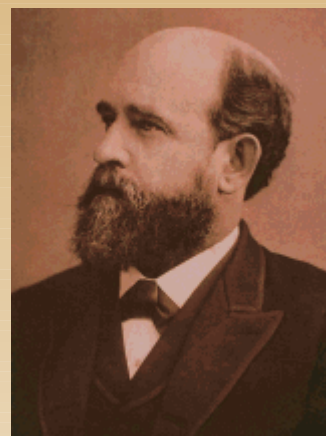
Now consider the incidence of a tax on land. As we first saw in Chapter 6, the burden of a tax falls more heavily on the side of the market that is less elastic. A tax on land takes this principle to an extreme. Because the elasticity of supply is zero, the landowners bear the entire burden of the tax.

Is there an ideal tax? Henry George, the nineteenth-century American economist and social philosopher, thought so. In his 1879 book *Progress and Poverty*, George argued that the government should raise all its revenue from a tax on land. This "single tax" was, he claimed, both equitable and efficient. George's ideas won him a large political following, and in 1886 he lost a close race for

Consider next the question of efficiency. As we just discussed, the deadweight loss of a tax depends on the elasticities of supply and demand. Again, a tax on land is an extreme case. Because supply is perfectly inelastic, a tax on land does not alter the market allocation. There is no deadweight loss, and the government's tax revenue exactly equals the loss of the landowners.

Although taxing land may look attractive in theory, it is not as straightforward in practice as it may appear. For a tax on land not to distort economic incentives, it must be a tax on raw land. Yet the value of land often comes from improvements, such as clearing trees, providing sewers, and building roads. Unlike the supply of raw land, the supply of improvements has an elasticity greater than zero. If a land tax were imposed on improvements, it would distort incentives. Landowners would respond by devoting fewer resources to improving their land.

Today, few economists support George's proposal for a single tax on land. Not only is taxing improvements a potential problem, but the tax would not raise enough revenue to pay for the much larger government we have today. Yet many of George's arguments remain valid. Here is the assessment of the eminent economist Milton Friedman a century after George's book: "In my opinion, the least bad tax is the property tax on the unimproved value of land, the Henry George argument of many, many years ago."

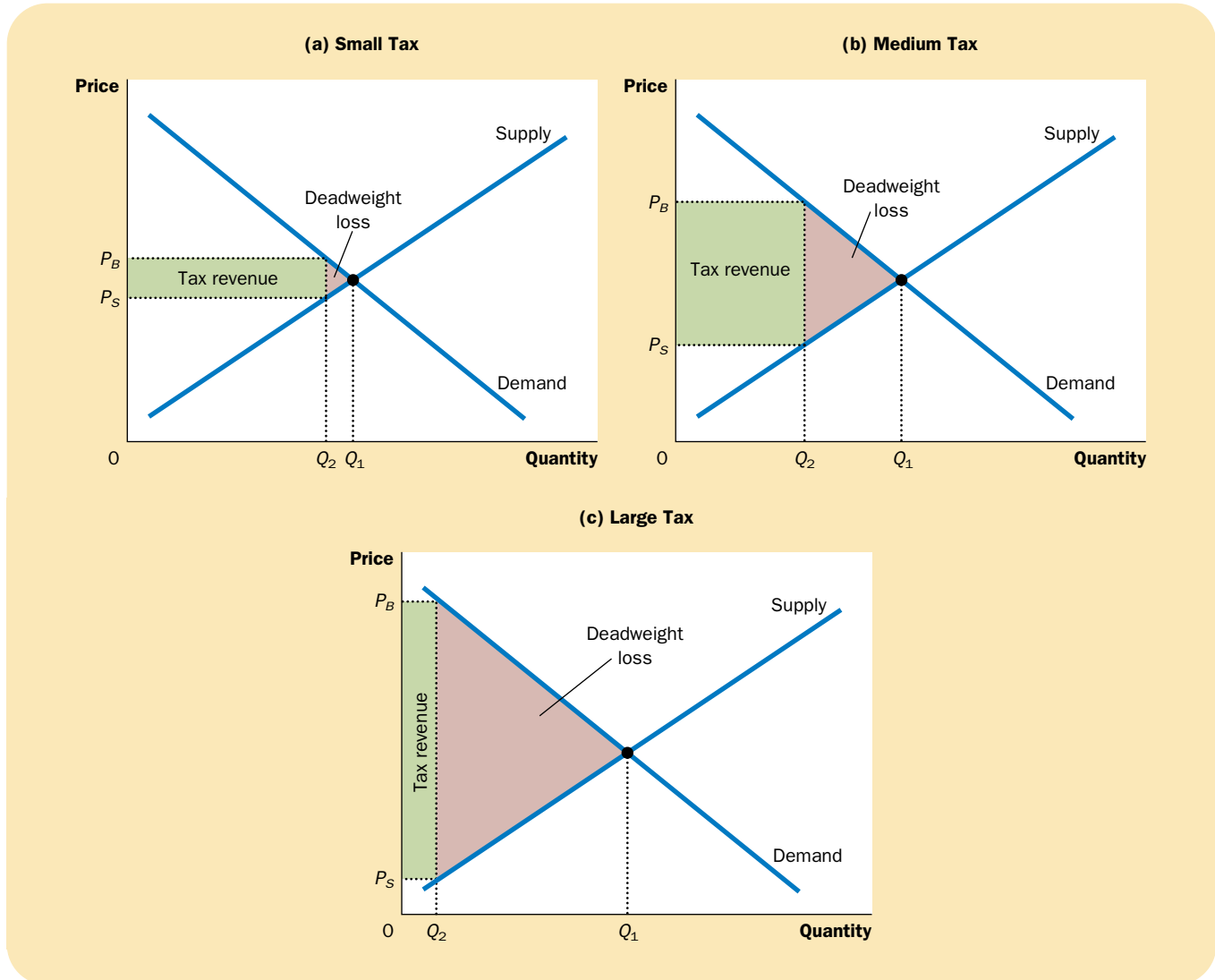


HENRY GEORGE

DEADWEIGHT LOSS AND
TAX REVENUE AS TAXES VARY

Taxes rarely stay the same for long periods of time. Policymakers in local, state, and federal governments are always considering raising one tax or lowering another. Here we consider what happens to the deadweight loss and tax revenue when the size of a tax changes.

Figure 8-6 shows the effects of a small, medium, and large tax, holding constant the market's supply and demand curves. The deadweight loss—the reduction in total surplus that results when the tax reduces the size of a market below



DEADWEIGHT LOSS AND TAX REVENUE FROM THREE TAXES OF DIFFERENT SIZE. The deadweight loss is the reduction in total surplus due to the tax. Tax revenue is the amount of the tax times the amount of the good sold. In panel (a), a small tax has a small deadweight loss and raises a small amount of revenue. In panel (b), a somewhat larger tax has a larger deadweight loss and raises a larger amount of revenue. In panel (c), a very large tax has a very large deadweight loss, but because it has reduced the size of the market so much, the tax raises only a small amount of revenue.

Figure 8-6

the optimum—equals the area of the triangle between the supply and demand curves. For the small tax in panel (a), the area of the deadweight loss triangle is quite small. But as the size of a tax rises in panels (b) and (c), the deadweight loss grows larger and larger.

Indeed, the deadweight loss of a tax rises even more rapidly than the size of the tax. The reason is that the deadweight loss is an area of a triangle, and an area

of a triangle depends on the *square* of its size. If we double the size of a tax, for instance, the base and height of the triangle double, so the deadweight loss rises by a factor of 4. If we triple the size of a tax, the base and height triple, so the deadweight loss rises by a factor of 9.

The government's tax revenue is the size of the tax times the amount of the good sold. As Figure 8-6 shows, tax revenue equals the area of the rectangle between the supply and demand curves. For the small tax in panel (a), tax revenue is small. As the size of a tax rises from panel (a) to panel (b), tax revenue grows. But as the size of the tax rises further from panel (b) to panel (c), tax revenue falls because the higher tax drastically reduces the size of the market. For a very large tax, no revenue would be raised, because people would stop buying and selling the good altogether.

Figure 8-7 summarizes these results. In panel (a) we see that as the size of a tax increases, its deadweight loss quickly gets larger. By contrast, panel (b) shows that tax revenue first rises with the size of the tax; but then, as the tax gets larger, the market shrinks so much that tax revenue starts to fall.

CASE STUDY THE LAFFER CURVE AND SUPPLY-SIDE ECONOMICS

One day in 1974, economist Arthur Laffer sat in a Washington restaurant with some prominent journalists and politicians. He took out a napkin and drew a figure on it to show how tax rates affect tax revenue. It looked much like panel (b) of our Figure 8-7. Laffer then suggested that the United States was on the downward-sloping side of this curve. Tax rates were so high, he argued, that reducing them would actually raise tax revenue.

Most economists were skeptical of Laffer's suggestion. The idea that a cut in tax rates could raise tax revenue was correct as a matter of economic theory, but there was more doubt about whether it would do so in practice. There was little evidence for Laffer's view that U.S. tax rates had in fact reached such extreme levels.

Nonetheless, the *Laffer curve* (as it became known) captured the imagination of Ronald Reagan. David Stockman, budget director in the first Reagan administration, offers the following story:

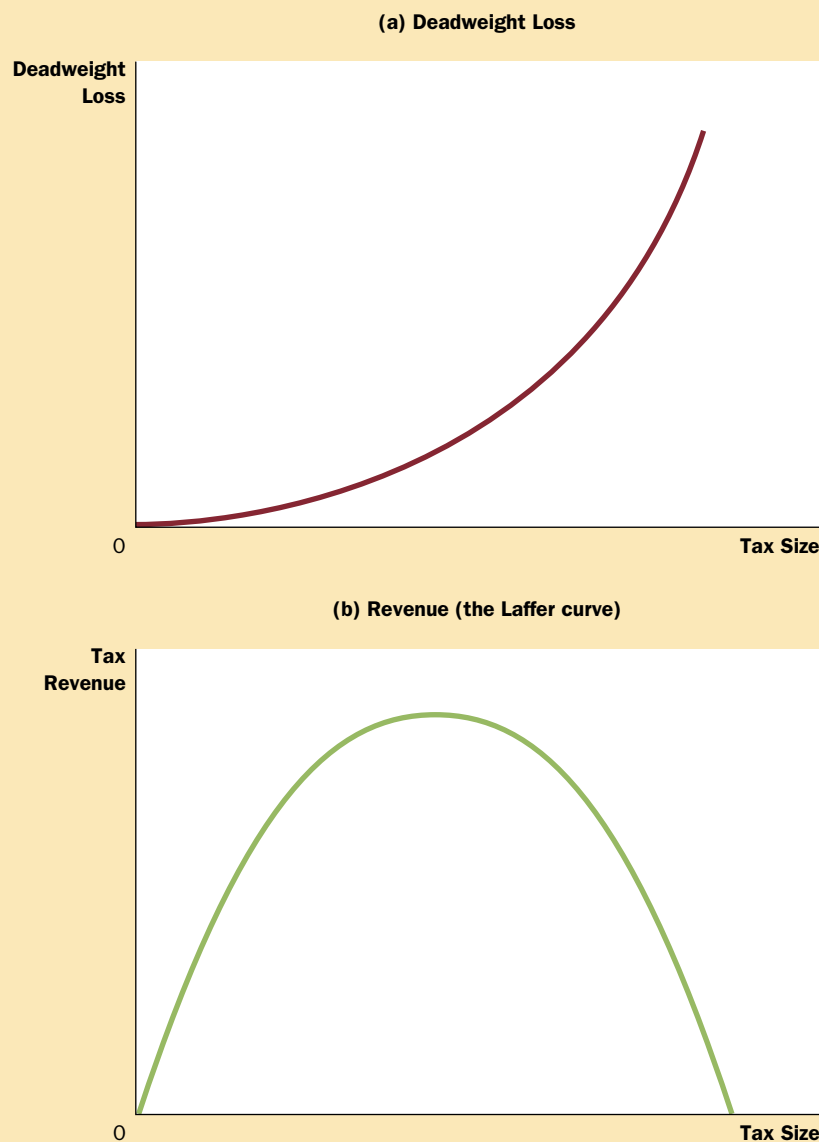
[Reagan] had once been on the Laffer curve himself. "I came into the Big Money making pictures during World War II," he would always say. At that time the wartime income surtax hit 90 percent. "You could only make four pictures and then you were in the top bracket," he would continue. "So we all quit working after four pictures and went off to the country." High tax rates caused less work. Low tax rates caused more. His experience proved it.

When Reagan ran for president in 1980, he made cutting taxes part of his platform. Reagan argued that taxes were so high that they were discouraging hard work. He argued that lower taxes would give people the proper incentive to work, which would raise economic well-being and perhaps even tax revenue. Because the cut in tax rates was intended to encourage people to increase the quantity of labor they supplied, the views of Laffer and Reagan became known as *supply-side economics*.

Subsequent history failed to confirm Laffer's conjecture that lower tax rates would raise tax revenue. When Reagan cut taxes after he was elected, the result

Figure 8-7

HOW DEADWEIGHT LOSS AND TAX REVENUE VARY WITH THE SIZE OF A TAX. Panel (a) shows that as the size of a tax grows larger, the deadweight loss grows larger. Panel (b) shows that tax revenue first rises, then falls. This relationship is sometimes called the Laffer curve.



was less tax revenue, not more. Revenue from personal income taxes (per person, adjusted for inflation) fell by 9 percent from 1980 to 1984, even though average income (per person, adjusted for inflation) grew by 4 percent over this period. The tax cut, together with policymakers' unwillingness to restrain spending, began a long period during which the government spent more than it collected in taxes. Throughout Reagan's two terms in office, and for many years thereafter, the government ran large budget deficits.

Yet Laffer's argument is not completely without merit. Although an overall cut in tax rates normally reduces revenue, some taxpayers at some times may be on the wrong side of the Laffer curve. In the 1980s, tax revenue collected from the richest Americans, who face the highest tax rates, did rise when their taxes were cut. The idea that cutting taxes can raise revenue may be correct if applied to

IN THE NEWS*How to Be Master
of the Universe*

WORLD LEADERS NEED TO UNDERSTAND the costs of taxation, even if the world they're leading happens to be the figment of some game designer's imagination.

Supply-Side Is a Winning Strategy

BY JOHN J. VECCHIONE

Congress may have given up on cutting taxes, but there's one corner of the country where supply-side economics still rules—the computer screens of game enthusiasts.

Not all messages from computer games are antisocial ones. Although we've heard a lot recently about games like Doom, known as "shooters," in what are known as "God games," a player assumes total control of a city, a country, or even a galaxy, deciding everything from

military to economic policy. In SimCity, a player runs a beleaguered municipal administration. In Civilization and its sequels, the player is the leader of a historic empire, such as Stalinist Russia or Elizabethan England, in a scramble for world domination. In Master of Orion, a player is given command of an entire species—whether humans or lizard-like Sakkras—with the goal of conquering the galaxy.

One thing these games have in common: Success requires economic growth, and that can only be achieved by keeping taxes low. Tax rates range from the edenic zero to the punitive 80%. With the proceeds of these taxes the player must build costly military or police forces and the infrastructure to support economic and technological advancement.

Why not simply keep taxes high and meet all the "societal needs" a despot could want? Because . . . keeping taxes high leads the population to produce less. As tax rates increase there is, at first, no easily discernable effect on the populace, except perhaps a few frowns and grumbles. But as soon as taxes reach a certain point—10% in some games, 20% in others—citizens begin to revolt. . . .

In games covering a single city, citizens vote with their feet and begin leaving town. No new jobs are created, and

once-vibrant downtown areas are left with little traffic but plenty of crime. Tax rates that approach 50% or more accelerate the trend. . . .

In the state or galaxy games, similar rules apply. During times of great military conflict or bursts of government construction, tax rates can be increased for a number of years without too much damage to the populace, and revenues do increase from the previous year. The government can simply buy what it needs from increased revenue. But a long war or government building program creates problems in "growing the economy" if tax rates are too high. Production slumps. The busy empire builder finds that his starships are harder to produce. Before long a once mighty empire is tottering on the brink of collapse and the ruler is deposed. The wise ruler keeps taxes as low as possible consistent with enough guns and roads to keep the country safe from a takeover by the enemy. . . .

Who says kids are wasting their time playing computers games?

SOURCE: *The Wall Street Journal*, May 5, 1999, p. A22.

those taxpayers facing the highest tax rates. In addition, Laffer's argument may be more plausible when applied to other countries, where tax rates are much higher than in the United States. In Sweden in the early 1980s, for instance, the typical worker faced a marginal tax rate of about 80 percent. Such a high tax rate provides a substantial disincentive to work. Studies have suggested that Sweden would indeed have raised more tax revenue if it had lowered its tax rates.

These ideas arise frequently in political debate. When Bill Clinton moved into the White House in 1993, he increased the federal income tax rates on high-income taxpayers to about 40 percent. Some economists criticized the policy, arguing that the plan would not yield as much revenue as the Clinton administration estimated. They claimed that the administration did not fully take into

account how taxes alter behavior. Conversely, when Bob Dole challenged Bill Clinton in the election of 1996, Dole proposed cutting personal income taxes. Although Dole rejected the idea that tax cuts would completely pay for themselves, he did claim that 28 percent of the tax cut would be recouped because lower tax rates would lead to more rapid economic growth. Economists debated whether Dole's 28 percent projection was reasonable, excessively optimistic, or (as Laffer might suggest) excessively pessimistic.

Policymakers disagree about these issues in part because they disagree about the size of the relevant elasticities. The more elastic that supply and demand are in any market, the more taxes in that market distort behavior, and the more likely it is that a tax cut will raise tax revenue. There is no debate, however, about the general lesson: How much revenue the government gains or loses from a tax change cannot be computed just by looking at tax rates. It also depends on how the tax change affects people's behavior.

QUICK QUIZ: If the government doubles the tax on gasoline, can you be sure that revenue from the gasoline tax will rise? Can you be sure that the deadweight loss from the gasoline tax will rise? Explain.

CONCLUSION

Taxes, Oliver Wendell Holmes once said, are the price we pay for a civilized society. Indeed, our society cannot exist without some form of taxes. We all expect the government to provide certain services, such as roads, parks, police, and national defense. These public services require tax revenue.

This chapter has shed some light on how high the price of civilized society can be. One of the *Ten Principles of Economics* discussed in Chapter 1 is that markets are usually a good way to organize economic activity. When the government imposes taxes on buyers or sellers of a good, however, society loses some of the benefits of market efficiency. Taxes are costly to market participants not only because taxes transfer resources from those participants to the government, but also because they alter incentives and distort market outcomes.



Summary

- ◆ A tax on a good reduces the welfare of buyers and sellers of the good, and the reduction in consumer and producer surplus usually exceeds the revenue raised by the government. The fall in total surplus—the sum of consumer surplus, producer surplus, and tax revenue—is called the deadweight loss of the tax.
- ◆ Taxes have deadweight losses because they cause buyers to consume less and sellers to produce less, and this change in behavior shrinks the size of the market below the level that maximizes total surplus. Because the elasticities of supply and demand measure how much market participants respond to market conditions, larger elasticities imply larger deadweight losses.
- ◆ As a tax grows larger, it distorts incentives more, and its deadweight loss grows larger. Tax revenue first rises with the size of a tax. Eventually, however, a larger tax reduces tax revenue because it reduces the size of the market.

Key Concepts

deadweight loss, p. 165

Questions for Review

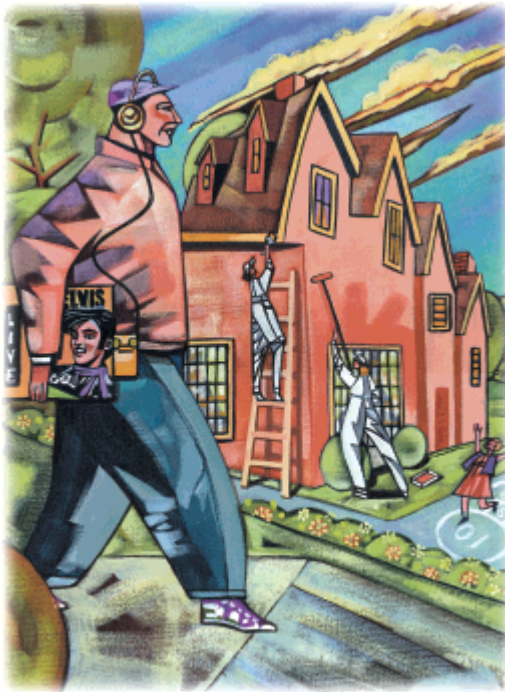
1. What happens to consumer and producer surplus when the sale of a good is taxed? How does the change in consumer and producer surplus compare to the tax revenue? Explain.
2. Draw a supply-and-demand diagram with a tax on the sale of the good. Show the deadweight loss. Show the tax revenue.
3. How do the elasticities of supply and demand affect the deadweight loss of a tax? Why do they have this effect?
4. Why do experts disagree about whether labor taxes have small or large deadweight losses?
5. What happens to the deadweight loss and tax revenue when a tax is increased?

Problems and Applications

1. The market for pizza is characterized by a downward-sloping demand curve and an upward-sloping supply curve.
 - a. Draw the competitive market equilibrium. Label the price, quantity, consumer surplus, and producer surplus. Is there any deadweight loss? Explain.
 - b. Suppose that the government forces each pizzeria to pay a \$1 tax on each pizza sold. Illustrate the effect of this tax on the pizza market, being sure to label the consumer surplus, producer surplus, government revenue, and deadweight loss. How does each area compare to the pre-tax case?
 - c. If the tax were removed, pizza eaters and sellers would be better off, but the government would lose tax revenue. Suppose that consumers and producers voluntarily transferred some of their gains to the government. Could all parties (including the government) be better off than they were with a tax? Explain using the labeled areas in your graph.
2. Evaluate the following two statements. Do you agree? Why or why not?
 - a. "If the government taxes land, wealthy land-owners will pass the tax on to their poorer renters."
 - b. "If the government taxes apartment buildings, wealthy landlords will pass the tax on to their poorer renters."
3. Evaluate the following two statements. Do you agree? Why or why not?
 - a. "A tax that has no deadweight loss cannot raise any revenue for the government."
 - b. "A tax that raises no revenue for the government cannot have any deadweight loss."
4. Consider the market for rubber bands.
 - a. If this market has very elastic supply and very inelastic demand, how would the burden of a tax on rubber bands be shared between consumers and producers? Use the tools of consumer surplus and producer surplus in your answer.
 - b. If this market has very inelastic supply and very elastic demand, how would the burden of a tax on rubber bands be shared between consumers and producers? Contrast your answer with your answer to part (a).
5. Suppose that the government imposes a tax on heating oil.
 - a. Would the deadweight loss from this tax likely be greater in the first year after it is imposed or in the fifth year? Explain.
 - b. Would the revenue collected from this tax likely be greater in the first year after it is imposed or in the fifth year? Explain.
6. After economics class one day, your friend suggests that taxing food would be a good way to raise revenue because the demand for food is quite inelastic. In what sense is taxing food a "good" way to raise revenue? In what sense is it not a "good" way to raise revenue?

7. Senator Daniel Patrick Moynihan once introduced a bill that would levy a 10,000 percent tax on certain hollow-tipped bullets.
- Do you expect that this tax would raise much revenue? Why or why not?
 - Even if the tax would raise no revenue, what might be Senator Moynihan's reason for proposing it?
8. The government places a tax on the purchase of socks.
- Illustrate the effect of this tax on equilibrium price and quantity in the sock market. Identify the following areas both before and after the imposition of the tax: total spending by consumers, total revenue for producers, and government tax revenue.
 - Does the price received by producers rise or fall? Can you tell whether total receipts for producers rise or fall? Explain.
 - Does the price paid by consumers rise or fall? Can you tell whether total spending by consumers rises or falls? Explain carefully. (Hint: Think about elasticity.) If total consumer spending falls, does consumer surplus rise? Explain.
9. Suppose the government currently raises \$100 million through a \$0.01 tax on widgets, and another \$100 million through a \$0.10 tax on gadgets. If the government doubled the tax rate on widgets and eliminated the tax on gadgets, would it raise more money than today, less money, or the same amount of money? Explain.
10. Most states tax the purchase of new cars. Suppose that New Jersey currently requires car dealers to pay the state \$100 for each car sold, and plans to increase the tax to \$150 per car next year.
- Illustrate the effect of this tax increase on the quantity of cars sold in New Jersey, the price paid by consumers, and the price received by producers.
 - Create a table that shows the levels of consumer surplus, producer surplus, government revenue, and total surplus both before and after the tax increase.
 - What is the change in government revenue? Is it positive or negative?
 - What is the change in deadweight loss? Is it positive or negative?
 - Give one reason why the demand for cars in New Jersey might be fairly elastic. Does this make the additional tax more or less likely to increase government revenue? How might states try to reduce the elasticity of demand?
11. Several years ago the British government imposed a "poll tax" that required each person to pay a flat amount to the government independent of his or her income or wealth. What is the effect of such a tax on economic efficiency? What is the effect on economic equity? Do you think this was a popular tax?
12. This chapter analyzed the welfare effects of a tax on a good. Consider now the opposite policy. Suppose that the government subsidizes a good: For each unit of the good sold, the government pays \$2 to the buyer. How does the subsidy affect consumer surplus, producer surplus, tax revenue, and total surplus? Does a subsidy lead to a deadweight loss? Explain.
13. (This problem uses some high school algebra and is challenging.) Suppose that a market is described by the following supply and demand equations:
- $$Q^S = 2P$$
- $$Q^D = 300 - P$$
- Solve for the equilibrium price and the equilibrium quantity.
 - Suppose that a tax of T is placed on buyers, so the new demand equation is
- $$Q^D = 300 - (P + T).$$
- Solve for the new equilibrium. What happens to the price received by sellers, the price paid by buyers, and the quantity sold?
- Tax revenue is $T \times Q$. Use your answer to part (b) to solve for tax revenue as a function of T . Graph this relationship for T between 0 and 300.
 - The deadweight loss of a tax is the area of the triangle between the supply and demand curves. Recalling that the area of a triangle is $1/2 \times \text{base} \times \text{height}$, solve for deadweight loss as a function of T . Graph this relationship for T between 0 and 300. (Hint: Looking sideways, the base of the deadweight loss triangle is T , and the height is the difference between the quantity sold with the tax and the quantity sold without the tax.)
 - The government now levies a tax on this good of \$200 per unit. Is this a good policy? Why or why not? Can you propose a better policy?

9



APPLICATION: INTERNATIONAL TRADE

If you check the labels on the clothes you are now wearing, you will probably find that some of your clothes were made in another country. A century ago the textiles and clothing industry was a major part of the U.S. economy, but that is no longer the case. Faced with foreign competitors that could produce quality goods at low cost, many U.S. firms found it increasingly difficult to produce and sell textiles and clothing at a profit. As a result, they laid off their workers and shut down their factories. Today, much of the textiles and clothing that Americans consume are imported from abroad.

The story of the textiles industry raises important questions for economic policy: How does international trade affect economic well-being? Who gains and who loses from free trade among countries, and how do the gains compare to the losses?

IN THIS CHAPTER YOU WILL . . .

Consider what determines whether a country imports or exports a good

Examine who wins and who loses from international trade

Learn that the gains to winners from international trade exceed the losses to losers

Analyze the welfare effects of tariffs and import quotas

Examine the arguments people use to advocate trade restrictions

Chapter 3 introduced the study of international trade by applying the principle of comparative advantage. According to this principle, all countries can benefit from trading with one another because trade allows each country to specialize in doing what it does best. But the analysis in Chapter 3 was incomplete. It did not explain how the international marketplace achieves these gains from trade or how the gains are distributed among various economic actors.

We now return to the study of international trade and take up these questions. Over the past several chapters, we have developed many tools for analyzing how markets work: supply, demand, equilibrium, consumer surplus, producer surplus, and so on. With these tools we can learn more about the effects of international trade on economic well-being.

THE DETERMINANTS OF TRADE

Consider the market for steel. The steel market is well suited to examining the gains and losses from international trade: Steel is made in many countries around the world, and there is much world trade in steel. Moreover, the steel market is one in which policymakers often consider (and sometimes implement) trade restrictions in order to protect domestic steel producers from foreign competitors. We examine here the steel market in the imaginary country of Isoland.

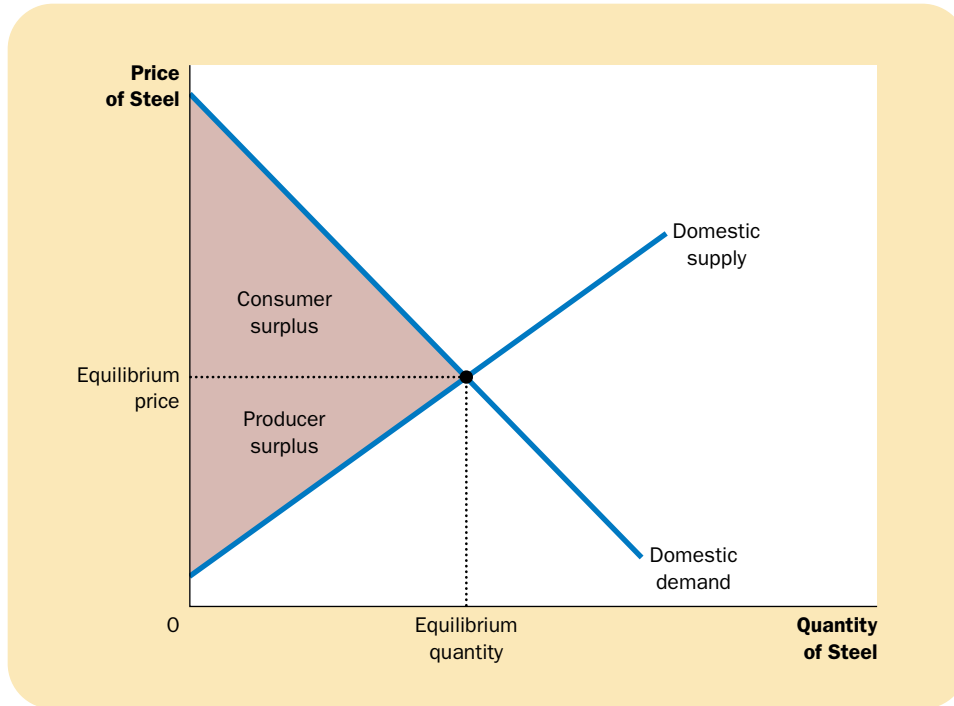
THE EQUILIBRIUM WITHOUT TRADE

As our story begins, the Isolandian steel market is isolated from the rest of the world. By government decree, no one in Isoland is allowed to import or export steel, and the penalty for violating the decree is so large that no one dares try.

Because there is no international trade, the market for steel in Isoland consists solely of Isolandian buyers and sellers. As Figure 9-1 shows, the domestic price adjusts to balance the quantity supplied by domestic sellers and the quantity demanded by domestic buyers. The figure shows the consumer and producer surplus in the equilibrium without trade. The sum of consumer and producer surplus measures the total benefits that buyers and sellers receive from the steel market.

Now suppose that, in an election upset, Isoland elects a new president. The president campaigned on a platform of “change” and promised the voters bold new ideas. Her first act is to assemble a team of economists to evaluate Isolandian trade policy. She asks them to report back on three questions:

- ◆ If the government allowed Isolandians to import and export steel, what would happen to the price of steel and the quantity of steel sold in the domestic steel market?
- ◆ Who would gain from free trade in steel and who would lose, and would the gains exceed the losses?
- ◆ Should a tariff (a tax on steel imports) or an import quota (a limit on steel imports) be part of the new trade policy?

**Figure 9-1**

THE EQUILIBRIUM WITHOUT INTERNATIONAL TRADE. When an economy cannot trade in world markets, the price adjusts to balance domestic supply and demand. This figure shows consumer and producer surplus in an equilibrium without international trade for the steel market in the imaginary country of Isoland.

After reviewing supply and demand in their favorite textbook (this one, of course), the Isolandian economics team begins its analysis.

THE WORLD PRICE AND COMPARATIVE ADVANTAGE

The first issue our economists take up is whether Isoland is likely to become a steel importer or a steel exporter. In other words, if free trade were allowed, would Isolandians end up buying or selling steel in world markets?

To answer this question, the economists compare the current Isolandian price of steel to the price of steel in other countries. We call the price prevailing in world markets the **world price**. If the world price of steel is higher than the domestic price, then Isoland would become an exporter of steel once trade is permitted. Isolandian steel producers would be eager to receive the higher prices available abroad and would start selling their steel to buyers in other countries. Conversely, if the world price of steel is lower than the domestic price, then Isoland would become an importer of steel. Because foreign sellers offer a better price, Isolandian steel consumers would quickly start buying steel from other countries.

In essence, comparing the world price and the domestic price before trade indicates whether Isoland has a comparative advantage in producing steel. The domestic price reflects the opportunity cost of steel: It tells us how much an Isolandian must give up to get one unit of steel. If the domestic price is low, the cost of producing steel in Isoland is low, suggesting that Isoland has a comparative advantage in producing steel relative to the rest of the world. If the domestic price is high, then the cost of producing steel in Isoland is high, suggesting that foreign countries have a comparative advantage in producing steel.

world price

the price of a good that prevails in the world market for that good

As we saw in Chapter 3, trade among nations is ultimately based on comparative advantage. That is, trade is beneficial because it allows each nation to specialize in doing what it does best. By comparing the world price and the domestic price before trade, we can determine whether Isoland is better or worse at producing steel than the rest of the world.

QUICK QUIZ: The country Autarka does not allow international trade. In Autarka, you can buy a wool suit for 3 ounces of gold. Meanwhile, in neighboring countries, you can buy the same suit for 2 ounces of gold. If Autarka were to allow free trade, would it import or export suits?

THE WINNERS AND LOSERS FROM TRADE

To analyze the welfare effects of free trade, the Isolandian economists begin with the assumption that Isoland is a small economy compared to the rest of the world so that its actions have negligible effect on world markets. The small-economy assumption has a specific implication for analyzing the steel market: If Isoland is a small economy, then the change in Isoland's trade policy will not affect the world price of steel. The Isolandians are said to be *price takers* in the world economy. That is, they take the world price of steel as given. They can sell steel at this price and be exporters or buy steel at this price and be importers.

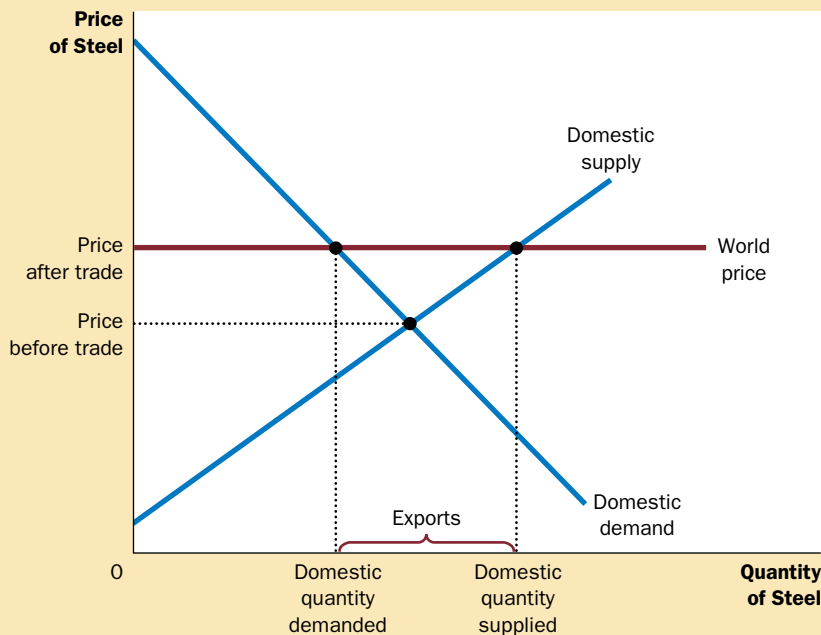
The small-economy assumption is not necessary to analyze the gains and losses from international trade. But the Isolandian economists know from experience that this assumption greatly simplifies the analysis. They also know that the basic lessons do not change in the more complicated case of a large economy.

THE GAINS AND LOSSES OF AN EXPORTING COUNTRY

Figure 9-2 shows the Isolandian steel market when the domestic equilibrium price before trade is below the world price. Once free trade is allowed, the domestic price rises to equal the world price. No seller of steel would accept less than the world price, and no buyer would pay more than the world price.

With the domestic price now equal to the world price, the domestic quantity supplied differs from the domestic quantity demanded. The supply curve shows the quantity of steel supplied by Isolandian sellers. The demand curve shows the quantity of steel demanded by Isolandian buyers. Because the domestic quantity supplied is greater than the domestic quantity demanded, Isoland sells steel to other countries. Thus, Isoland becomes a steel exporter.

Although domestic quantity supplied and domestic quantity demanded differ, the steel market is still in equilibrium because there is now another participant in the market: the rest of the world. One can view the horizontal line at the world price as representing the demand for steel from the rest of the world. This demand curve is perfectly elastic because Isoland, as a small economy, can sell as much steel as it wants at the world price.

Figure 9-2

INTERNATIONAL TRADE IN AN EXPORTING COUNTRY. Once trade is allowed, the domestic price rises to equal the world price. The supply curve shows the quantity of steel produced domestically, and the demand curve shows the quantity consumed domestically. Exports from Isoland equal the difference between the domestic quantity supplied and the domestic quantity demanded at the world price.

Now consider the gains and losses from opening up trade. Clearly, not everyone benefits. Trade forces the domestic price to rise to the world price. Domestic producers of steel are better off because they can now sell steel at a higher price, but domestic consumers of steel are worse off because they have to buy steel at a higher price.

To measure these gains and losses, we look at the changes in consumer and producer surplus, which are shown in Figure 9-3 and summarized in Table 9-1. Before trade is allowed, the price of steel adjusts to balance domestic supply and domestic demand. Consumer surplus, the area between the demand curve and the before-trade price, is area $A + B$. Producer surplus, the area between the supply curve and the before-trade price, is area C . Total surplus before trade, the sum of consumer and producer surplus, is area $A + B + C$.

After trade is allowed, the domestic price rises to the world price. Consumer surplus is area A (the area between the demand curve and the world price). Producer surplus is area $B + C + D$ (the area between the supply curve and the world price). Thus, total surplus with trade is area $A + B + C + D$.

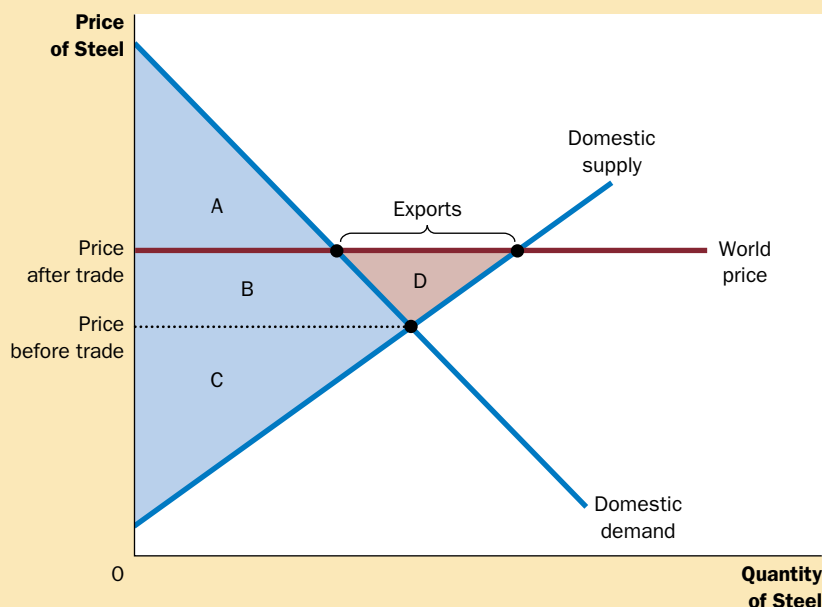
These welfare calculations show who wins and who loses from trade in an exporting country. Sellers benefit because producer surplus increases by the area $B + D$. Buyers are worse off because consumer surplus decreases by the area B . Because the gains of sellers exceed the losses of buyers by the area D , total surplus in Isoland increases.

This analysis of an exporting country yields two conclusions:

- ◆ When a country allows trade and becomes an exporter of a good, domestic producers of the good are better off, and domestic consumers of the good are worse off.

Figure 9-3

HOW FREE TRADE AFFECTS WELFARE IN AN EXPORTING COUNTRY. When the domestic price rises to equal the world price, sellers are better off (producer surplus rises from C to $B + C + D$), and buyers are worse off (consumer surplus falls from $A + B$ to A). Total surplus rises by an amount equal to area D , indicating that trade raises the economic well-being of the country as a whole.

**Table 9-1**

CHANGES IN WELFARE FROM FREE TRADE: THE CASE OF AN EXPORTING COUNTRY. The table examines changes in economic welfare resulting from opening up a market to international trade. Letters refer to the regions marked in Figure 9-3.

	BEFORE TRADE	AFTER TRADE	CHANGE
Consumer Surplus	$A + B$	A	$-B$
Producer Surplus	C	$B + C + D$	$+(B + D)$
Total Surplus	$A + B + C$	$A + B + C + D$	$+D$

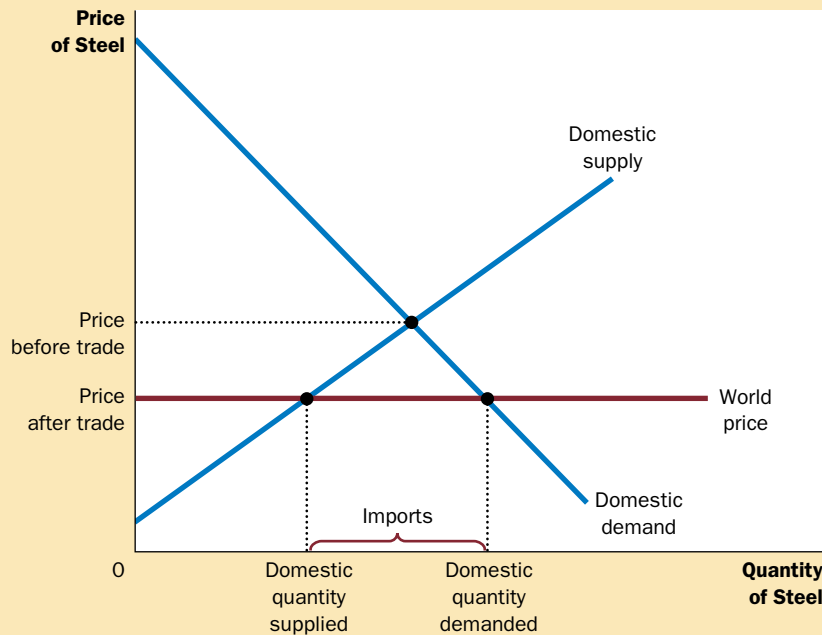
The area D shows the increase in total surplus and represents the gains from trade.

- ◆ Trade raises the economic well-being of a nation in the sense that the gains of the winners exceed the losses of the losers.

THE GAINS AND LOSSES OF AN IMPORTING COUNTRY

Now suppose that the domestic price before trade is above the world price. Once again, after free trade is allowed, the domestic price must equal the world price. As Figure 9-4 shows, the domestic quantity supplied is less than the domestic quantity demanded. The difference between the domestic quantity demanded and the domestic quantity supplied is bought from other countries, and Isoland becomes a steel importer.

In this case, the horizontal line at the world price represents the supply of the rest of the world. This supply curve is perfectly elastic because Isoland is a small economy and, therefore, can buy as much steel as it wants at the world price.

Figure 9-4

INTERNATIONAL TRADE IN AN IMPORTING COUNTRY. Once trade is allowed, the domestic price falls to equal the world price. The supply curve shows the amount produced domestically, and the demand curve shows the amount consumed domestically. Imports equal the difference between the domestic quantity demanded and the domestic quantity supplied at the world price.

Now consider the gains and losses from trade. Once again, not everyone benefits. When trade forces the domestic price to fall, domestic consumers are better off (they can now buy steel at a lower price), and domestic producers are worse off (they now have to sell steel at a lower price). Changes in consumer and producer surplus measure the size of the gains and losses, as shown in Figure 9-5 and Table 9-2. Before trade, consumer surplus is area A, producer surplus is area B + C, and total surplus is area A + B + C. After trade is allowed, consumer surplus is area A + B + D, producer surplus is area C, and total surplus is area A + B + C + D.

These welfare calculations show who wins and who loses from trade in an importing country. Buyers benefit because consumer surplus increases by the area B + D. Sellers are worse off because producer surplus falls by the area B. The gains of buyers exceed the losses of sellers, and total surplus increases by the area D.

This analysis of an importing country yields two conclusions parallel to those for an exporting country:

- ◆ When a country allows trade and becomes an importer of a good, domestic consumers of the good are better off, and domestic producers of the good are worse off.
- ◆ Trade raises the economic well-being of a nation in the sense that the gains of the winners exceed the losses of the losers.

Now that we have completed our analysis of trade, we can better understand one of the *Ten Principles of Economics* in Chapter 1: Trade can make everyone better off. If Isoland opens up its steel market to international trade, that change will create



Figure 9-5

HOW FREE TRADE AFFECTS WELFARE IN AN IMPORTING COUNTRY. When the domestic price falls to equal the world price, buyers are better off (consumer surplus rises from A to $A + B + D$), and sellers are worse off (producer surplus falls from $B + C$ to C). Total surplus rises by an amount equal to area D , indicating that trade raises the economic well-being of the country as a whole.

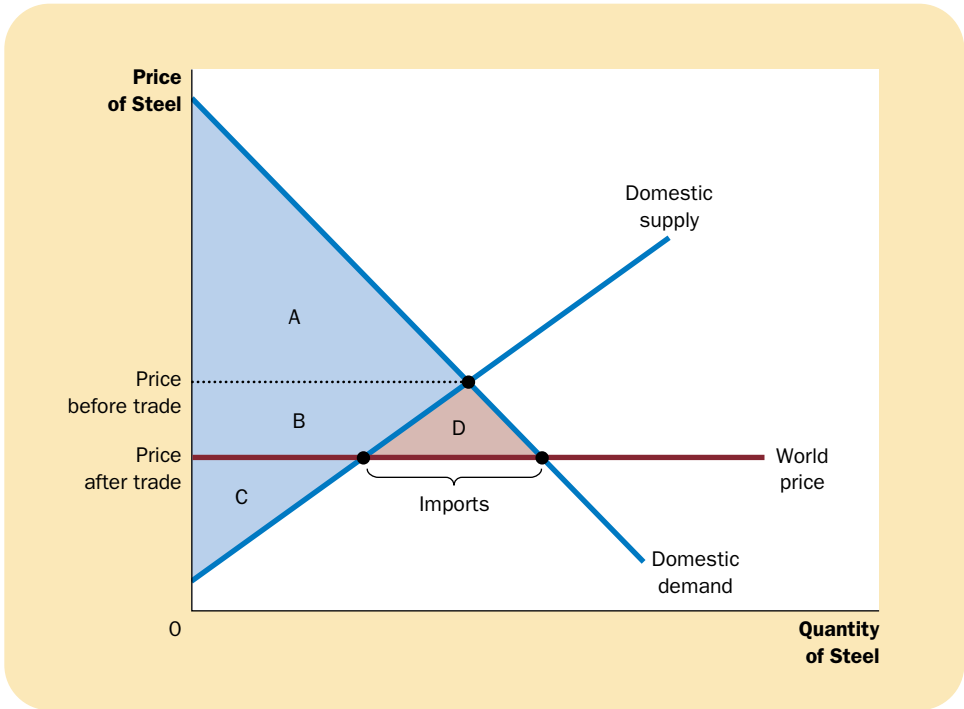


Table 9-2

CHANGES IN WELFARE FROM FREE TRADE: THE CASE OF AN IMPORTING COUNTRY. The table examines changes in economic welfare resulting from opening up a market to international trade. Letters refer to the regions marked in Figure 9-5.

	BEFORE TRADE	AFTER TRADE	CHANGE
Consumer Surplus	A	$A + B + D$	$+(B + D)$
Producer Surplus	$B + C$	C	$-B$
Total Surplus	$A + B + C$	$A + B + C + D$	$+D$

The area D shows the increase in total surplus and represents the gains from trade.

winners and losers, regardless of whether Isoland ends up exporting or importing steel. In either case, however, the gains of the winners exceed the losses of the losers, so the winners could compensate the losers and still be better off. In this sense, trade *can* make everyone better off. But *will* trade make everyone better off? Probably not. In practice, compensation for the losers from international trade is rare. Without such compensation, opening up to international trade is a policy that expands the size of the economic pie, while perhaps leaving some participants in the economy with a smaller slice.

THE EFFECTS OF A TARIFF

tariff

a tax on goods produced abroad and sold domestically

The Isolandian economists next consider the effects of a **tariff**—a tax on imported goods. The economists quickly realize that a tariff on steel will have no effect if Isoland becomes a steel exporter. If no one in Isoland is interested in importing

IN THE NEWS

Life in Isoland

OUR STORY ABOUT THE STEEL INDUSTRY and the debate over trade policy in Isoland is just a parable. Or is it?

Clinton Warns U.S. Will Fight Cheap Imports

BY DAVID E. SANGER

President Clinton said for the first time today that the United States would not tolerate the "flooding of our markets" with low-cost goods from Asia and Russia, particularly steel, that are threatening the jobs of American workers.

The President's statement came days after a White House meeting of top executives of steel companies and the United Steelworkers of America, which helped get out the vote for Democrats last week, playing a pivotal role with other unions in the party's success in midterm elections.

After the meeting, which included Mr. Clinton, Vice President Al Gore, and top Cabinet officials, aides said the White House would not grant the unions' demand to cut off imports of steel they say are being dumped in the American markets. But today, the President warned that foreign nations must "play by the rules," appearing to signal that the United States would press other nations to restrict their exports to the United States. [Author's note: In the end, the Clinton administration did decide to limit steel imports.]

SOURCE: *The New York Times*, November 11, 1998, p A1.

steel, a tax on steel imports is irrelevant. The tariff matters only if Isoland becomes a steel importer. Concentrating their attention on this case, the economists compare welfare with and without the tariff.

Figure 9-6 shows the Isolandian market for steel. Under free trade, the domestic price equals the world price. A tariff raises the price of imported steel above the world price by the amount of the tariff. Domestic suppliers of steel, who compete with suppliers of imported steel, can now sell their steel for the world price plus the amount of the tariff. Thus, the price of steel—both imported and domestic—rises by the amount of the tariff and is, therefore, closer to the price that would prevail without trade.

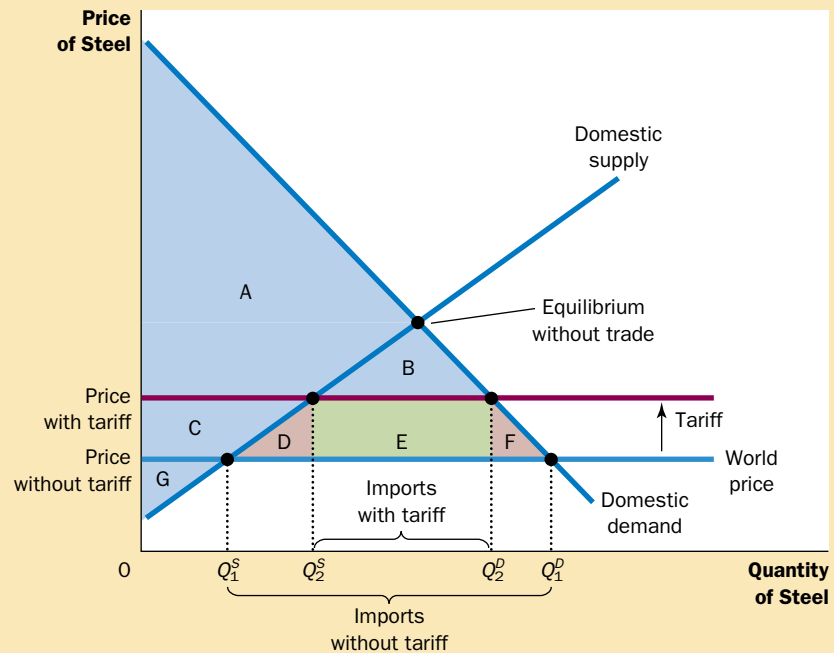
The change in price affects the behavior of domestic buyers and sellers. Because the tariff raises the price of steel, it reduces the domestic quantity demanded from Q_1^D to Q_2^D and raises the domestic quantity supplied from Q_1^S to Q_2^S . Thus, the tariff reduces the quantity of imports and moves the domestic market closer to its equilibrium without trade.

Now consider the gains and losses from the tariff. Because the tariff raises the domestic price, domestic sellers are better off, and domestic buyers are worse off. In addition, the government raises revenue. To measure these gains and losses, we look at the changes in consumer surplus, producer surplus, and government revenue. These changes are summarized in Table 9-3.

Before the tariff, the domestic price equals the world price. Consumer surplus, the area between the demand curve and the world price, is area $A + B + C + D + E + F$. Producer surplus, the area between the supply curve and the world price, is area G . Government revenue equals zero. Total surplus, the sum of consumer surplus, producer surplus, and government revenue, is area $A + B + C + D + E + F + G$.

Figure 9-6

THE EFFECTS OF A TARIFF. A tariff reduces the quantity of imports and moves a market closer to the equilibrium that would exist without trade. Total surplus falls by an amount equal to area $D + F$. These two triangles represent the deadweight loss from the tariff.



	BEFORE TARIFF	AFTER TARIFF	CHANGE
Consumer Surplus	$A + B + C + D + E + F$	$A + B$	$-(C + D + E + F)$
Producer Surplus	G	$C + G$	$+C$
Government Revenue	None	E	$+E$
Total Surplus	$A + B + C + D + E + F + G$	$A + B + C + E + G$	$-(D + F)$

The area $D + F$ shows the fall in total surplus and represents the deadweight loss of the tariff.

Table 9-3

CHANGES IN WELFARE FROM A TARIFF. The table compares economic welfare when trade is unrestricted and when trade is restricted with a tariff. Letters refer to the regions marked in Figure 9-6.

Once the government imposes a tariff, the domestic price exceeds the world price by the amount of the tariff. Consumer surplus is now area $A + B$. Producer surplus is area $C + G$. Government revenue, which is the quantity of after-tariff imports times the size of the tariff, is the area E . Thus, total surplus with the tariff is area $A + B + C + E + G$.

To determine the total welfare effects of the tariff, we add the change in consumer surplus (which is negative), the change in producer surplus (positive), and the change in government revenue (positive). We find that total surplus in the market decreases by the area $D + F$. This fall in total surplus is called the *deadweight loss* of the tariff.

It is not surprising that a tariff causes a deadweight loss, because a tariff is a type of tax. Like any tax on the sale of a good, it distorts incentives and pushes the allocation of scarce resources away from the optimum. In this case, we can identify two effects. First, the tariff on steel raises the price of steel that domestic producers can charge above the world price and, as a result, encourages them to increase production of steel (from Q_1^S to Q_2^S). Second, the tariff raises the price that domestic steel buyers have to pay and, therefore, encourages them to reduce consumption of steel (from Q_1^D to Q_2^D). Area D represents the deadweight loss from the overproduction of steel, and area F represents the deadweight loss from the underconsumption. The total deadweight loss of the tariff is the sum of these two triangles.

THE EFFECTS OF AN IMPORT QUOTA

The Isolandian economists next consider the effects of an **import quota**—a limit on the quantity of imports. In particular, imagine that the Isolandian government distributes a limited number of import licenses. Each license gives the license holder the right to import 1 ton of steel into Isoland from abroad. The Isolandian economists want to compare welfare under a policy of free trade and welfare with the addition of this import quota.

import quota

a limit on the quantity of a good that can be produced abroad and sold domestically

Figure 9-7 shows how an import quota affects the Isolandian market for steel. Because the import quota prevents Isolandians from buying as much steel as they want from abroad, the supply of steel is no longer perfectly elastic at the world price. Instead, as long as the price of steel in Isoland is above the world price, the license holders import as much as they are permitted, and the total supply of steel in Isoland equals the domestic supply plus the quota amount. That is, the supply curve above the world price is shifted to the right by exactly the amount of the quota. (The supply curve below the world price does not shift because, in this case, importing is not profitable for the license holders.)

The price of steel in Isoland adjusts to balance supply (domestic plus imported) and demand. As the figure shows, the quota causes the price of steel to rise above the world price. The domestic quantity demanded falls from Q_1^D to Q_2^D , and the domestic quantity supplied rises from Q_1^S to Q_2^S . Not surprisingly, the import quota reduces steel imports.

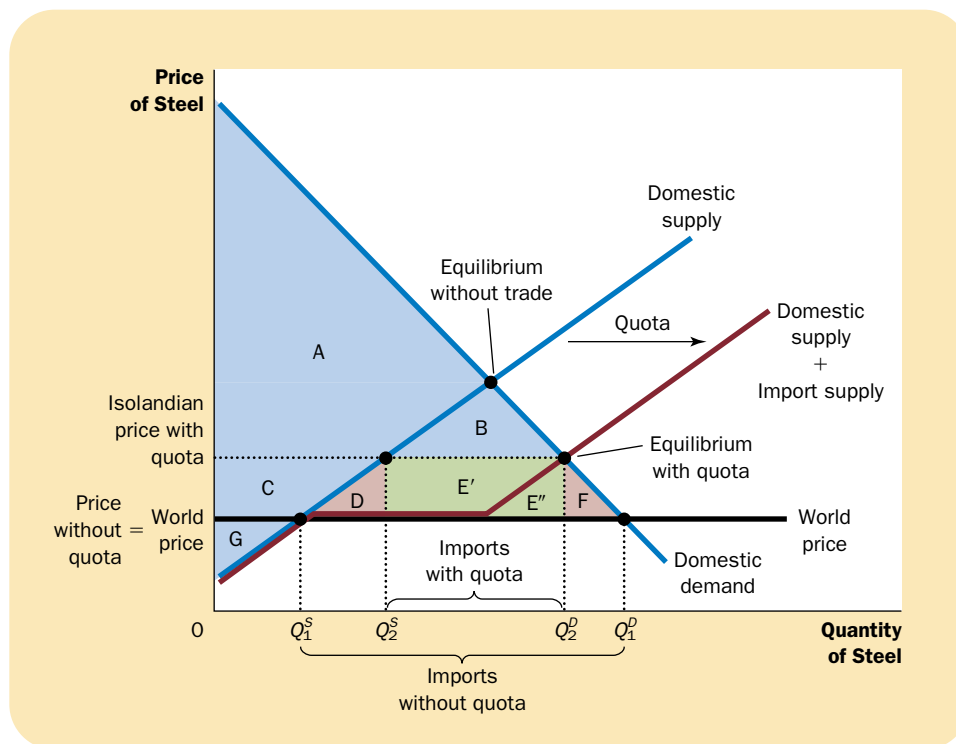
Now consider the gains and losses from the quota. Because the quota raises the domestic price above the world price, domestic sellers are better off, and domestic buyers are worse off. In addition, the license holders are better off because they make a profit from buying at the world price and selling at the higher domestic price. To measure these gains and losses, we look at the changes in consumer surplus, producer surplus, and license-holder surplus, as shown in Table 9-4.

Before the government imposes the quota, the domestic price equals the world price. Consumer surplus, the area between the demand curve and the world price, is area $A + B + C + D + E' + E'' + F$. Producer surplus, the area between the supply curve and the world price, is area G . The surplus of license holders equals zero because there are no licenses. Total surplus, the sum of consumer, producer, and license-holder surplus, is area $A + B + C + D + E' + E'' + F + G$.

After the government imposes the import quota and issues the licenses, the domestic price exceeds the world price. Domestic consumers get surplus equal to area $A + B$, and domestic producers get surplus equal to area $C + G$. The license holders make a profit on each unit imported equal to the difference between the

Figure 9-7

THE EFFECTS OF AN IMPORT QUOTA. An import quota, like a tariff, reduces the quantity of imports and moves a market closer to the equilibrium that would exist without trade. Total surplus falls by an amount equal to area $D + F$. These two triangles represent the deadweight loss from the quota. In addition, the import quota transfers $E' + E''$ to whoever holds the import licenses.



	BEFORE QUOTA	AFTER QUOTA	CHANGE
Consumer Surplus	$A + B + C + D + E' + E'' + F$	$A + B$	$-(C + D + E' + E'' + F)$
Producer Surplus	G	$C + G$	$+C$
License-Holder Surplus	None	$E' + E''$	$+(E' + E'')$
Total Surplus	$A + B + C + D + E' + E'' + F + G$	$A + B + C + E' + E'' + G$	$-(D + F)$

The area $D + F$ shows the fall in total surplus and represents the deadweight loss of the quota.

Table 9-4

CHANGES IN WELFARE FROM AN IMPORT QUOTA. The table compares economic welfare when trade is unrestricted and when trade is restricted with an import quota. Letters refer to the regions marked in Figure 9-7.

Isolandinian price of steel and the world price. Their surplus equals this price differential times the quantity of imports. Thus, it equals the area of the rectangle $E' + E''$. Total surplus with the quota is the area $A + B + C + E' + E'' + G$.

To see how total welfare changes with the imposition of the quota, we add the change in consumer surplus (which is negative), the change in producer surplus (positive), and the change in license-holder surplus (positive). We find that total surplus in the market decreases by the area $D + F$. This area represents the deadweight loss of the import quota.

This analysis should seem somewhat familiar. Indeed, if you compare the analysis of import quotas in Figure 9-7 with the analysis of tariffs in Figure 9-6, you will see that they are essentially identical. *Both tariffs and import quotas raise the domestic price of the good, reduce the welfare of domestic consumers, increase the welfare of domestic producers, and cause deadweight losses.* There is only one difference between these two types of trade restriction: A tariff raises revenue for the government (area E in Figure 9-6), whereas an import quota creates surplus for license holders (area E' + E'' in Figure 9-7).

Tariffs and import quotas can be made to look even more similar. Suppose that the government tries to capture the license-holder surplus for itself by charging a fee for the licenses. A license to sell 1 ton of steel is worth exactly the difference between the Isolandian price of steel and the world price, and the government can set the license fee as high as this price differential. If the government does this, the license fee for imports works exactly like a tariff: Consumer surplus, producer surplus, and government revenue are exactly the same under the two policies.

In practice, however, countries that restrict trade with import quotas rarely do so by selling the import licenses. For example, the U.S. government has at times pressured Japan to “voluntarily” limit the sale of Japanese cars in the United States. In this case, the Japanese government allocates the import licenses to Japanese firms, and the surplus from these licenses (area E' + E'') accrues to those firms. This kind of import quota is, from the standpoint of U.S. welfare, strictly worse than a U.S. tariff on imported cars. Both a tariff and an import quota raise prices, restrict trade, and cause deadweight losses, but at least the tariff produces revenue for the U.S. government rather than for Japanese auto companies.

Although in our analysis so far import quotas and tariffs appear to cause similar deadweight losses, a quota can potentially cause an even larger deadweight loss, depending on the mechanism used to allocate the import licenses. Suppose that when Isoland imposes a quota, everyone understands that the licenses will go to those who spend the most resources lobbying the Isolandian government. In this case, there is an implicit license fee—the cost of lobbying. The revenues from this fee, however, rather than being collected by the government, are spent on lobbying expenses. The deadweight losses from this type of quota include not only the losses from overproduction (area D) and underconsumption (area F) but also whatever part of the license-holder surplus (area E' + E'') is wasted on the cost of lobbying.

THE LESSONS FOR TRADE POLICY

The team of Isolandian economists can now write to the new president:

Dear Madam President,

You asked us three questions about opening up trade. After much hard work, we have the answers.

Question: If the government allowed Isolandians to import and export steel, what would happen to the price of steel and the quantity of steel sold in the domestic steel market?

Answer: Once trade is allowed, the Isolandian price of steel would be driven to equal the price prevailing around the world.

If the world price is now higher than the Isolandian price, our price would rise. The higher price would reduce the amount of steel Isolandians consume and raise the amount of steel that Isolandians produce. Isoland would, therefore, become a steel exporter. This occurs because, in this case, Isoland would have a comparative advantage in producing steel.

Conversely, if the world price is now lower than the Isolandian price, our price would fall. The lower price would raise the amount of steel that Isolandians consume and lower the amount of steel that Isolandians produce. Isoland would, therefore, become a steel importer. This occurs because, in this case, other countries would have a comparative advantage in producing steel.

Question: Who would gain from free trade in steel and who would lose, and would the gains exceed the losses?

Answer: The answer depends on whether the price rises or falls when trade is allowed. If the price rises, producers of steel gain, and consumers of steel lose. If the price falls, consumers gain, and producers lose. In both cases, the gains are larger than the losses. Thus, free trade raises the total welfare of Isolandians.

Question: Should a tariff or an import quota be part of the new trade policy?

Answer: A tariff, like most taxes, has deadweight losses: The revenue raised would be smaller than the losses to the buyers and sellers. In this case, the deadweight losses occur because the tariff would move the economy closer to our current no-trade equilibrium. An import quota works much like a tariff and would cause similar deadweight losses. The best policy, from the standpoint of economic efficiency, would be to allow trade without a tariff or an import quota.

We hope you find these answers helpful as you decide on your new policy.

Your faithful servants,
Isolandian economics team

QUICK QUIZ: Draw the supply and demand curve for wool suits in the country of Autarka. When trade is allowed, the price of a suit falls from 3 to 2 ounces of gold. In your diagram, what is the change in consumer surplus, the change in producer surplus, and the change in total surplus? How would a tariff on suit imports alter these effects?

THE ARGUMENTS FOR RESTRICTING TRADE

The letter from the economics team persuades the new president of Isoland to consider opening up trade in steel. She notes that the domestic price is now high compared to the world price. Free trade would, therefore, cause the price of steel to fall and hurt domestic steel producers. Before implementing the new policy, she asks Isolandian steel companies to comment on the economists' advice.

FYI

Other Benefits
of International
Trade

Our conclusions so far have been based on the standard analysis of international trade. As we have seen, there are winners and losers when a nation opens itself up to trade, but the gains to the winners exceed the losses of the losers. Yet the case for free trade can be made even stronger. There are several other economic benefits of trade beyond those emphasized in the standard analysis.

Here, in a nutshell, are some of these other benefits:

- ◆ *Increased variety of goods:* Goods produced in different countries are not exactly the same. German beer, for instance, is not the same as American beer. Free trade gives consumers in all countries greater variety from which to choose.
- ◆ *Lower costs through economies of scale:* Some goods can be produced at low cost only if they are produced in large quantities—a phenomenon called *economies of scale*. A firm in a small country cannot take full advantage of economies of scale if it can sell only in a small domestic market. Free trade gives firms access to larger world markets and allows them to realize economies of scale more fully.

- ◆ *Increased competition:* A company shielded from foreign competitors is more likely to have market power, which in turn gives it the ability to raise prices above competitive levels. This is a type of market failure. Opening up trade fosters competition and gives the invisible hand a better chance to work its magic.
- ◆ *Enhanced flow of ideas:* The transfer of technological advances around the world is often thought to be linked to international trade in the goods that embody those advances. The best way for a poor, agricultural nation to learn about the computer revolution, for instance, is to buy some computers from abroad, rather than trying to make them domestically.

Thus, free international trade increases variety for consumers, allows firms to take advantage of economies of scale, makes markets more competitive, and facilitates the spread of technology. If the Isolandian economists thought these effects were important, their advice to their president would be even more forceful.

Not surprisingly, the steel companies are opposed to free trade in steel. They believe that the government should protect the domestic steel industry from foreign competition. Let's consider some of the arguments they might give to support their position and how the economics team would respond.

THE JOBS ARGUMENT

Opponents of free trade often argue that trade with other countries destroys domestic jobs. In our example, free trade in steel would cause the price of steel to fall, reducing the quantity of steel produced in Isoland and thus reducing employment in the Isolandian steel industry. Some Isolandian steelworkers would lose their jobs.

Yet free trade creates jobs at the same time that it destroys them. When Isolandians buy steel from other countries, those countries obtain the resources to buy other goods from Isoland. Isolandian workers would move from the steel industry to those industries in which Isoland has a comparative advantage. Although the transition may impose hardship on some workers in the short run, it allows Isolandians as a whole to enjoy a higher standard of living.

Opponents of trade are often skeptical that trade creates jobs. They might respond that *everything* can be produced more cheaply abroad. Under free trade, they might argue, Isolandians could not be profitably employed in any industry.

Berry's World



"You like protectionism as a 'working man.' How about as a consumer?"

As Chapter 3 explains, however, the gains from trade are based on comparative advantage, not absolute advantage. Even if one country is better than another country at producing everything, each country can still gain from trading with the other. Workers in each country will eventually find jobs in the industry in which that country has a comparative advantage.

THE NATIONAL-SECURITY ARGUMENT

When an industry is threatened with competition from other countries, opponents of free trade often argue that the industry is vital for national security. In our example, Isolandian steel companies might point out that steel is used to make guns and tanks. Free trade would allow Isoland to become dependent on foreign countries to supply steel. If a war later broke out, Isoland might be unable to produce enough steel and weapons to defend itself.

Economists acknowledge that protecting key industries may be appropriate when there are legitimate concerns over national security. Yet they fear that this argument may be used too quickly by producers eager to gain at consumers' expense. The U.S. watchmaking industry, for instance, long argued that it was vital for national security, claiming that its skilled workers would be necessary in wartime. Certainly, it is tempting for those in an industry to exaggerate their role in national defense in order to obtain protection from foreign competition.

THE INFANT-INDUSTRY ARGUMENT

New industries sometimes argue for temporary trade restrictions to help them get started. After a period of protection, the argument goes, these industries will mature and be able to compete with foreign competitors. Similarly, older industries sometimes argue that they need temporary protection to help them adjust to new conditions. For example, General Motors Chairman Roger Smith once argued for temporary protection "to give U.S. automakers turnaround time to get the domestic industry back on its feet."

Economists are often skeptical about such claims. The primary reason is that the infant-industry argument is difficult to implement in practice. To apply protection successfully, the government would need to decide which industries will eventually be profitable and decide whether the benefits of establishing these industries exceed the costs to consumers of protection. Yet "picking winners" is extraordinarily difficult. It is made even more difficult by the political process, which often awards protection to those industries that are politically powerful. And once a powerful industry is protected from foreign competition, the "temporary" policy is hard to remove.

In addition, many economists are skeptical about the infant-industry argument even in principle. Suppose, for instance, that the Isolandian steel industry is young and unable to compete profitably against foreign rivals. Yet there is reason to believe that the industry can be profitable in the long run. In this case, the owners of the firms should be willing to incur temporary losses in order to obtain the eventual profits. Protection is not necessary for an industry to grow. Firms in various industries—such as many Internet firms today—incur temporary losses in the hope of growing and becoming profitable in the future. And many of them succeed, even without protection from foreign competition.

THE UNFAIR-COMPETITION ARGUMENT

A common argument is that free trade is desirable only if all countries play by the same rules. If firms in different countries are subject to different laws and regulations, then it is unfair (the argument goes) to expect the firms to compete in the international marketplace. For instance, suppose that the government of Neighborland subsidizes its steel industry by giving steel companies large tax breaks. The Isolandian steel industry might argue that it should be protected from this foreign competition because Neighborland is not competing fairly.

Would it, in fact, hurt Isoland to buy steel from another country at a subsidized price? Certainly, Isolandian steel producers would suffer, but Isolandian steel consumers would benefit from the low price. Moreover, the case for free trade is no different: The gains of the consumers from buying at the low price would exceed the losses of the producers. Neighborland's subsidy to its steel industry may be a bad policy, but it is the taxpayers of Neighborland who bear the burden. Isoland can benefit from the opportunity to buy steel at a subsidized price.

THE PROTECTION-AS-A-BARGAINING-CHIP ARGUMENT

Another argument for trade restrictions concerns the strategy of bargaining. Many policymakers claim to support free trade but, at the same time, argue that trade restrictions can be useful when we bargain with our trading partners. They claim that the threat of a trade restriction can help remove a trade restriction already imposed by a foreign government. For example, Isoland might threaten to impose a tariff on steel unless Neighborland removes its tariff on wheat. If Neighborland responds to this threat by removing its tariff, the result can be freer trade.

The problem with this bargaining strategy is that the threat may not work. If it doesn't work, the country has a difficult choice. It can carry out its threat and implement the trade restriction, which would reduce its own economic welfare. Or it can back down from its threat, which would cause it to lose prestige in international affairs. Faced with this choice, the country would probably wish that it had never made the threat in the first place.

An example of this occurred in 1999, when the U.S. government accused Europeans of restricting the import of U.S. bananas. After a long and bitter dispute with governments that are normally U.S. allies, the United States placed 100 percent tariffs on a range of European products from cheese to cashmere. In the end, not only were Europeans denied the benefits of American bananas, but Americans were denied the benefits of European cheese. Sometimes, when a government engages in a game of brinkmanship, as the United States did in this case, everyone goes over the brink together.

CASE STUDY TRADE AGREEMENTS

A country can take one of two approaches to achieving free trade. It can take a *unilateral* approach and remove its trade restrictions on its own. This is the approach that Great Britain took in the nineteenth century and that Chile and South Korea have taken in recent years. Alternatively, a country can take a *multilateral* approach and reduce its trade restrictions while other countries do the

IN THE NEWS

A Chicken Invasion

WHEN DOMESTIC PRODUCERS COMPLAIN about competition from abroad, they often assert that consumers are not well served by imperfect foreign products. The following article documents how Russian producers of chicken reacted to competition from the United States.

U.S. Chicken in Every Pot? Nyet! Russians Cry Foul

BY MICHAEL R. GORDON

Moscow—A nasty little skirmish between Russia and the United States is brewing here over a threatened trade barrier.

But this fight is not about manufactured consumer goods or high technology, but about American chicken, which has flooded the Russian market.

To the frustration, and considerable anxiety, of American companies, the Russian government has threatened to ban further American poultry sales effective March 19. . . .

The ostensible reason for the Russian government's warning is health—a

seemingly strange concern in a country with a generally lax record in observing safety standards, where virtually every able-bodied man and woman smokes.

Today, no less an authority than the Veterinary Department of the Russian Agriculture and Food Ministry said the ban was needed to protect consumers here against infected poultry until the United States improved its standards.

But the real agenda, American producers contend, is old-fashioned protectionism.

Agitated Russian producers, whose birds, Russian consumers say, are no match for their American competition in terms of quality and price, have repeatedly complained that the United States is trying to destroy the Russian poultry industry and capture its market. And now American companies fear the Russian producers are striking back. . . .

The first big invasion of frozen poultry [into Russia] came during the Bush administration. . . . The export proved to be very popular with Russian consumers, who dubbed them Bush legs.

After the demise of the Soviet Union, American poultry exports continued to soar. Russian poultry production, meanwhile, fell 40 percent, the result of rising grain prices and declining subsidies.

Astoundingly, a third of all American exports to Russia is poultry, American officials say. . . .

If the confrontation continues, the United States has a number of possible



A THREAT TO RUSSIA?

recourses, including arguing that the Russian action is inconsistent with Moscow's bid to join the World Trade Organization.

Some experts, however, believe there is an important countervailing force here that may lead to a softening of the Russian position: namely Russian consumers.

Russian consumers favor the American birds, which despite the dire warnings of the Russian government, have come to symbolize quality. And they vote, too.

SOURCE: *The New York Times*, February 24, 1996, pp. 33, 34.

same. In other words, it can bargain with its trading partners in an attempt to reduce trade restrictions around the world.

One important example of the multilateral approach is the North American Free Trade Agreement (NAFTA), which in 1993 lowered trade barriers among the United States, Mexico, and Canada. Another is the General Agreement on

Tariffs and Trade (GATT), which is a continuing series of negotiations among many of the world's countries with the goal of promoting free trade. The United States helped to found GATT after World War II in response to the high tariffs imposed during the Great Depression of the 1930s. Many economists believe that the high tariffs contributed to the economic hardship during that period. GATT has successfully reduced the average tariff among member countries from about 40 percent after World War II to about 5 percent today. The rules established under GATT are now enforced by an international institution called the World Trade Organization (WTO).

What are the pros and cons of the multilateral approach to free trade? One advantage is that the multilateral approach has the potential to result in freer trade than a unilateral approach because it can reduce trade restrictions abroad as well as at home. If international negotiations fail, however, the result could be more restricted trade than under a unilateral approach.

In addition, the multilateral approach may have a political advantage. In most markets, producers are fewer and better organized than consumers—and thus wield greater political influence. Reducing the Isolandian tariff on steel, for example, may be politically difficult if considered by itself. The steel companies would oppose free trade, and the users of steel who would benefit are so numerous that organizing their support would be difficult. Yet suppose that Neighborland promises to reduce its tariff on wheat at the same time that Island reduces its tariff on steel. In this case, the Isolandian wheat farmers, who are also politically powerful, would back the agreement. Thus, the multilateral approach to free trade can sometimes win political support when a unilateral reduction cannot.

QUICK QUIZ: The textile industry of Autarka advocates a ban on the import of wool suits. Describe five arguments its lobbyists might make. Give a response to each of these arguments.

CONCLUSION

Economists and the general public often disagree about free trade. In 1993, for example, the United States faced the question of whether to ratify the North American Free Trade Agreement, which reduced trade restrictions among the United States, Canada, and Mexico. Opinion polls showed the general public in the United States about evenly split on the issue, and the agreement passed in Congress by only a narrow margin. Opponents viewed free trade as a threat to job security and the American standard of living. By contrast, economists overwhelmingly supported the agreement. They viewed free trade as a way of allocating production efficiently and raising living standards in all three countries.

Economists view the United States as an ongoing experiment that confirms the virtues of free trade. Throughout its history, the United States has allowed unrestricted trade among the states, and the country as a whole has benefited from the specialization that trade allows. Florida grows oranges, Texas pumps oil, California makes wine, and so on. Americans would not enjoy the high standard of living

IN THE NEWS*The Case for
Unilateral Disarmament
in the Trade Wars*

ECONOMIST JAGDISH BHAGWATI ARGUES that the United States should lower its trade barriers unilaterally.

Free Trade without Treaties

BY JAGDISH BHAGWATI

President Clinton and 17 other Asian-Pacific leaders are meeting today in Vancouver. Rather than the convivial photo-op they'd planned, however, they must contend with worrisome trade news. A spate of Asian currency devaluations has raised the specter of renewed protectionism around the world. South America's Mercosur trade bloc, led by Brazil, just raised its tariffs some 30 percent. And Congress turned its back on the president and refused to approve fast-track authority for him to negotiate further free-trade accords. [Author's

note: Fast-track authority would allow the president to negotiate trade deals that Congress would consider without the ability to attach amendments.]

In light of all this dismaying news, what are the prospects for free trade? Is the future bleak, or will the postwar trend of dramatic liberalization continue to accelerate despite these setbacks?

The immediate prospects for more U.S.-led multilateral trade accords do indeed look grim after the defeat of fast-track. But that doesn't mean that free trade itself is on the ropes. A large portion of the world's trade liberalization in the last quarter-century has been *unilateral*. Those countries that lower trade barriers of their own accord not only profit themselves, but also often induce the laggards to match their example. The most potent force for the worldwide freeing of trade, then, is unilateral U.S. action. If the United States continues to do away with tariffs and trade barriers, other countries will follow suit—fast-track or no fast-track.

To be sure, the General Agreement on Tariffs and Trade, the World Trade Organization, and other multilateral tariff reductions have greatly contributed to global wealth. The WTO has become the international institution for setting the "rules" on public and private practices

that affect competition among trading nations. Much still needs to be done in that mode, particularly on agriculture tariffs, which remain too high around the world. A future U.S. president, if not Mr. Clinton, will certainly need fast-track authority if another multilateral effort, such as the "millennium round" called for by Sir Leon Brittan of the European Union, is to pursue these goals.

But the good news is that even if organized labor, radical environmentalists, and others who fear the global economy continue to impede fast-track during Congress's next session, they cannot stop the historic freeing of trade that has been occurring unilaterally worldwide.

From the 1970s through the 1990s, Latin America witnessed dramatic lowering of trade barriers unilaterally by Chile, Bolivia, and Paraguay; and the entire continent has been moving steadily toward further trade liberalization. Mercosur's recent actions are a setback, but only a small one—so far.

Latin America's record has been bettered by unilateral liberalizers in Asia and the Pacific. New Zealand began dismantling its substantial trade protection apparatus in 1985. That effort was driven by the reformist views of then-Prime Minister David Lange, who declared, "In

they do today if people could consume only those goods and services produced in their own states. The world could similarly benefit from free trade among countries.

To better understand economists' view of trade, let's continue our parable. Suppose that the country of Isoland ignores the advice of its economics team and decides not to allow free trade in steel. The country remains in the equilibrium without international trade.

Then, one day, some Isolandian inventor discovers a new way to make steel at very low cost. The process is quite mysterious, however, and the inventor insists on keeping it a secret. What is odd is that the inventor doesn't need any workers or iron ore to make steel. The only input he requires is wheat.

the course of about three years we changed from being a country run like a Polish shipyard into one that could be internationally competitive."

Since the 1980s, Hong Kong's and Singapore's enormous successes as free traders have served as potent examples of unilateral market opening, encouraging Indonesia, the Philippines, Thailand, South Korea, and Malaysia to follow suit. By 1991 even India, which has been astonishingly autarkic for more than four decades, had finally learned the virtue of free trade and had embarked on a massive lowering of its tariffs and non-tariff barriers.

In Central and Eastern Europe, the collapse of communism led to a wholesale, unilateral, and nondiscriminatory removal of trade barriers as well. The French economist Patrick Messerlin has shown how this happened in three waves: Czechoslovakia, Poland, and Hungary liberalized right after the fall of the Berlin Wall; next came Bulgaria, Romania, and Slovenia; and finally, the Baltic countries began unilateral opening in 1991. . . .

U.S. leadership is crucial to maintaining the trend toward free trade. Such ultramodern industries as telecommunications and financial services gained their momentum largely from unilateral

openness and deregulation in the United States. This in turn led to a softening of protectionist attitudes in the European Union and Japan.

These developed economies are now moving steadily in the direction of openness and competition—not because any officials in Washington threaten them with retribution, but because they've seen how U.S. companies become more competitive once regulation and other trade barriers have fallen. A Brussels bureaucrat can argue with a Washington bureaucrat, but he cannot argue with the markets. Faced with the prospect of being elbowed out of world markets by American firms, Japan and Europe have no option but to follow the U.S. example, belatedly but surely, in opening their own markets.

The biggest threat to free trade is not the loss of fast-track per se, but the signal it sends that Americans may not be interested in lowering their trade barriers any further. To counteract this attitude, President Clinton needs to mount the bully pulpit and explain the case for free trade—a case that Adam Smith first made more than 200 years ago, but that continues to come under attack.

The president, free from the burdens of constituency interests that cripple many in Congress, could argue,

credibly and with much evidence, that free trade is in the interest of the whole world, but that, because the U.S. economy is the most competitive anywhere, we have the most to gain. The president could also point to plenty of evidence that debunks the claims of protectionists. The unions may argue that trade with poor countries depresses our workers' wages, for example, but in fact the best evidence shows that such trade has *helped* workers by moderating the fall in their wages from technological changes.

Assuming that the president can make the case for free trade at home, the prospects for free trade worldwide remain bright. The United States doesn't need to sign treaties to open markets or, heaven forbid, issue counterproductive threats to close our own markets if others are less open than we are. We simply need to offer an example of openness and deregulation to the rest of the world. Other countries will see our success, and seek to emulate it.

SOURCE: *The Wall Street Journal*, November 24, 1997, p. A22.

The inventor is hailed as a genius. Because steel is used in so many products, the invention lowers the cost of many goods and allows all Isolandians to enjoy a higher standard of living. Workers who had previously produced steel do suffer when their factories close, but eventually they find work in other industries. Some become farmers and grow the wheat that the inventor turns into steel. Others enter new industries that emerge as a result of higher Isolandian living standards. Everyone understands that the displacement of these workers is an inevitable part of progress.

After several years, a newspaper reporter decides to investigate this mysterious new steel process. She sneaks into the inventor's factory and learns that the inventor is a fraud. The inventor has not been making steel at all. Instead, he has

been smuggling wheat abroad in exchange for steel from other countries. The only thing that the inventor had discovered was the gains from international trade.

When the truth is revealed, the government shuts down the inventor's operation. The price of steel rises, and workers return to jobs in steel factories. Living standards in Island fall back to their former levels. The inventor is jailed and held up to public ridicule. After all, he was no inventor. He was just an economist.

Summary

- ◆ The effects of free trade can be determined by comparing the domestic price without trade to the world price. A low domestic price indicates that the country has a comparative advantage in producing the good and that the country will become an exporter. A high domestic price indicates that the rest of the world has a comparative advantage in producing the good and that the country will become an importer.
- ◆ When a country allows trade and becomes an exporter of a good, producers of the good are better off, and consumers of the good are worse off. When a country allows trade and becomes an importer of a good, consumers are better off, and producers are worse off. In both cases, the gains from trade exceed the losses.
- ◆ A tariff—a tax on imports—moves a market closer to the equilibrium that would exist without trade and, therefore, reduces the gains from trade. Although domestic producers are better off and the government raises revenue, the losses to consumers exceed these gains.
- ◆ An import quota has effects that are similar to those of a tariff. Under a quota, however, the holders of the import licenses receive the revenue that the government would collect with a tariff.
- ◆ There are various arguments for restricting trade: protecting jobs, defending national security, helping infant industries, preventing unfair competition, and responding to foreign trade restrictions. Although some of these arguments have some merit in some cases, economists believe that free trade is usually the better policy.

Key Concepts

world price, p. 181

tariff, p. 186

import quota, p. 189

Questions for Review

1. What does the domestic price that prevails without international trade tell us about a nation's comparative advantage?
2. When does a country become an exporter of a good? An importer?
3. Draw the supply-and-demand diagram for an importing country. What is consumer surplus and producer surplus before trade is allowed? What is consumer surplus and producer surplus with free trade? What is the change in total surplus?
4. Describe what a tariff is, and describe its economic effects.
5. What is an import quota? Compare its economic effects with those of a tariff.
6. List five arguments often given to support trade restrictions. How do economists respond to these arguments?
7. What is the difference between the unilateral and multilateral approaches to achieving free trade? Give an example of each.

Problems and Applications

1. The United States represents a small part of the world orange market.
 - a. Draw a diagram depicting the equilibrium in the U.S. orange market without international trade. Identify the equilibrium price, equilibrium quantity, consumer surplus, and producer surplus.
 - b. Suppose that the world orange price is below the U.S. price before trade, and that the U.S. orange market is now opened to trade. Identify the new equilibrium price, quantity consumed, quantity produced domestically, and quantity imported. Also show the change in the surplus of domestic consumers and producers. Has domestic total surplus increased or decreased?
2. The world price of wine is below the price that would prevail in the United States in the absence of trade.
 - a. Assuming that American imports of wine are a small part of total world wine production, draw a graph for the U.S. market for wine under free trade. Identify consumer surplus, producer surplus, and total surplus in an appropriate table.
 - b. Now suppose that an unusual shift of the Gulf Stream leads to an unseasonably cold summer in Europe, destroying much of the grape harvest there. What effect does this shock have on the world price of wine? Using your graph and table from part (a), show the effect on consumer surplus, producer surplus, and total surplus in the United States. Who are the winners and losers? Is the United States as a whole better or worse off?
3. The world price of cotton is below the no-trade price in Country A and above the no-trade price in Country B. Using supply-and-demand diagrams and welfare tables such as those in the chapter, show the gains from trade in each country. Compare your results for the two countries.
4. Suppose that Congress imposes a tariff on imported autos to protect the U.S. auto industry from foreign competition. Assuming that the U.S. is a price taker in the world auto market, show on a diagram: the change in the quantity of imports, the loss to U.S. consumers, the gain to U.S. manufacturers, government revenue, and the deadweight loss associated with the tariff. The loss to consumers can be decomposed into three pieces: a transfer to domestic producers, a transfer to the government, and a deadweight loss. Use your diagram to identify these three pieces.
5. According to an article in *The New York Times* (Nov. 5, 1993), “many Midwest wheat farmers oppose the [North American] free trade agreement [NAFTA] as much as many corn farmers support it.” For simplicity, assume that the United States is a small country in the markets for both corn and wheat, and that without the free trade agreement, the United States would not trade these commodities internationally. (Both of these assumptions are false, but they do not affect the qualitative responses to the following questions.)
 - a. Based on this report, do you think the world wheat price is above or below the U.S. no-trade wheat price? Do you think the world corn price is above or below the U.S. no-trade corn price? Now analyze the welfare consequences of NAFTA in both markets.
 - b. Considering both markets together, does NAFTA make U.S. farmers as a group better or worse off? Does it make U.S. consumers as a group better or worse off? Does it make the United States as a whole better or worse off?
6. Imagine that winemakers in the state of Washington petitioned the state government to tax wines imported from California. They argue that this tax would both raise tax revenue for the state government and raise employment in the Washington state wine industry. Do you agree with these claims? Is it a good policy?
7. Senator Ernest Hollings once wrote that “consumers *do not* benefit from lower-priced imports. Glance through some mail-order catalogs and you’ll see that consumers pay exactly the same price for clothing whether it is U.S.-made or imported.” Comment.
8. Write a brief essay advocating or criticizing each of the following policy positions:
 - a. The government should not allow imports if foreign firms are selling below their costs of production (a phenomenon called “dumping”).
 - b. The government should temporarily stop the import of goods for which the domestic industry is new and struggling to survive.
 - c. The government should not allow imports from countries with weaker environmental regulations than ours.
9. Suppose that a technological advance in Japan lowers the world price of televisions.

- a. Assume the U.S. is an importer of televisions and there are no trade restrictions. How does the technological advance affect the welfare of U.S. consumers and U.S. producers? What happens to total surplus in the United States?
 - b. Now suppose the United States has a quota on television imports. How does the Japanese technological advance affect the welfare of U.S. consumers, U.S. producers, and the holders of import licenses?
10. When the government of Tradeland decides to impose an import quota on foreign cars, three proposals are suggested: (1) Sell the import licenses in an auction. (2) Distribute the licenses randomly in a lottery. (3) Let people wait in line and distribute the licenses on a first-come, first-served basis. Compare the effects of these policies. Which policy do you think has the largest deadweight losses? Which policy has the smallest deadweight losses? Why? (Hint: The government's other ways of raising tax revenue all cause deadweight losses themselves.)
11. An article in *The Wall Street Journal* (June 26, 1990) about sugar beet growers explained that "the government props up domestic sugar prices by curtailing imports of lower-cost sugar. Producers are guaranteed a 'market stabilization price' of \$0.22 a pound, about \$0.09 higher than the current world market price." The government maintains the higher price by imposing an import quota.
 - a. Illustrate the effect of this quota on the U.S. sugar market. Label the relevant prices and quantities under free trade and under the quota.
 - b. Analyze the effects of the sugar quota using the tools of welfare analysis.
 - c. The article also comments that "critics of the sugar program say that [the quota] has deprived numerous sugar-producing nations in the Caribbean, Latin America, and Far East of export earnings, harmed their economies, and caused political instability, while increasing Third World demand for U.S. foreign aid." Our usual welfare analysis includes only gains and losses to U.S. consumers and producers. What role do you think the gains or losses to people in other countries should play in our economic policymaking?
 - d. The article continues that "at home, the sugar program has helped make possible the spectacular rise of the high-fructose corn syrup industry." Why has the sugar program had this effect? (Hint: Are sugar and corn syrup substitutes or complements?)
12. (This question is challenging.) Consider a small country that exports steel. Suppose that a "pro-trade" government decides to subsidize the export of steel by paying a certain amount for each ton sold abroad. How does this export subsidy affect the domestic price of steel, the quantity of steel produced, the quantity of steel consumed, and the quantity of steel exported? How does it affect consumer surplus, producer surplus, government revenue, and total surplus? (Hint: The analysis of an export subsidy is similar to the analysis of a tariff.)

10



EXTERNALITIES

Firms that make and sell paper also create, as a by-product of the manufacturing process, a chemical called dioxin. Scientists believe that once dioxin enters the environment, it raises the population's risk of cancer, birth defects, and other health problems.

Is the production and release of dioxin a problem for society? In Chapters 4 through 9 we examined how markets allocate scarce resources with the forces of supply and demand, and we saw that the equilibrium of supply and demand is typically an efficient allocation of resources. To use Adam Smith's famous metaphor, the "invisible hand" of the marketplace leads self-interested buyers and sellers in a market to maximize the total benefit that society derives from that market. This insight is the basis for one of the *Ten Principles of Economics* in Chapter 1: Markets are usually a good way to organize economic activity. Should we conclude, therefore, that the invisible hand prevents firms in the paper market from emitting too much dioxin?

IN THIS CHAPTER YOU WILL . . .

***Learn the nature of
an externality***

***See why
externalities can
make market
outcomes inefficient***

***Examine how people
can sometimes solve
the problem of
externalities on
their own***

***Consider why private
solutions to
externalities
sometimes do not
work***

***Examine the various
government policies
aimed at solving the
problem of
externalities***

**externality**

the uncompensated impact of one person's actions on the well-being of a bystander

Markets do many things well, but they do not do everything well. In this chapter we begin our study of another of the *Ten Principles of Economics*: Governments can sometimes improve market outcomes. We examine why markets sometimes fail to allocate resources efficiently, how government policies can potentially improve the market's allocation, and what kinds of policies are likely to work best.

The market failures examined in this chapter fall under a general category called *externalities*. An **externality** arises when a person engages in an activity that influences the well-being of a bystander and yet neither pays nor receives any compensation for that effect. If the impact on the bystander is adverse, it is called a *negative externality*; if it is beneficial, it is called a *positive externality*. In the presence of externalities, society's interest in a market outcome extends beyond the well-being of buyers and sellers in the market; it also includes the well-being of bystanders who are affected. Because buyers and sellers neglect the external effects of their actions when deciding how much to demand or supply, the market equilibrium is not efficient when there are externalities. That is, the equilibrium fails to maximize the total benefit to society as a whole. The release of dioxin into the environment, for instance, is a negative externality. Self-interested paper firms will not consider the full cost of the pollution they create and, therefore, will emit too much pollution unless the government prevents or discourages them from doing so.

Externalities come in many varieties, as do the policy responses that try to deal with the market failure. Here are some examples:

- ◆ The exhaust from automobiles is a negative externality because it creates smog that other people have to breathe. As a result of this externality, drivers tend to pollute too much. The federal government attempts to solve this problem by setting emission standards for cars. It also taxes gasoline to reduce the amount that people drive.
- ◆ Restored historic buildings convey a positive externality because people who walk or ride by them can enjoy their beauty and the sense of history that these buildings provide. Building owners do not get the full benefit of restoration and, therefore, tend to discard older buildings too quickly. Many local governments respond to this problem by regulating the destruction of historic buildings and by providing tax breaks to owners who restore them.
- ◆ Barking dogs create a negative externality because neighbors are disturbed by the noise. Dog owners do not bear the full cost of the noise and, therefore, tend to take too few precautions to prevent their dogs from barking. Local governments address this problem by making it illegal to "disturb the peace."
- ◆ Research into new technologies provides a positive externality because it creates knowledge that other people can use. Because inventors cannot capture the full benefits of their inventions, they tend to devote too few resources to research. The federal government addresses this problem partially through the patent system, which gives inventors an exclusive use over their inventions for a period of time.

In each of these cases, some decisionmaker is failing to take account of the external effects of his or her behavior. The government responds by trying to influence this behavior to protect the interests of bystanders.

EXTERNALITIES AND MARKET INEFFICIENCY

In this section we use the tools from Chapter 7 to examine how externalities affect economic well-being. The analysis shows precisely why externalities cause markets to allocate resources inefficiently. Later in the chapter we examine various ways in which private actors and public policymakers may remedy this type of market failure.

WELFARE ECONOMICS: A RECAP

We begin by recalling the key lessons of welfare economics from Chapter 7. To make our analysis concrete, we will consider a specific market—the market for aluminum. Figure 10-1 shows the supply and demand curves in the market for aluminum.

As you should recall from Chapter 7, the supply and demand curves contain important information about costs and benefits. The demand curve for aluminum reflects the value of aluminum to consumers, as measured by the prices they are willing to pay. At any given quantity, the height of the demand curve shows the willingness to pay of the marginal buyer. In other words, it shows the value to the consumer of the last unit of aluminum bought. Similarly, the supply curve reflects the costs of producing aluminum. At any given quantity, the height of the supply curve shows the cost of the marginal seller. In other words, it shows the cost to the producer of the last unit of aluminum sold.

In the absence of government intervention, the price adjusts to balance the supply and demand for aluminum. The quantity produced and consumed in the

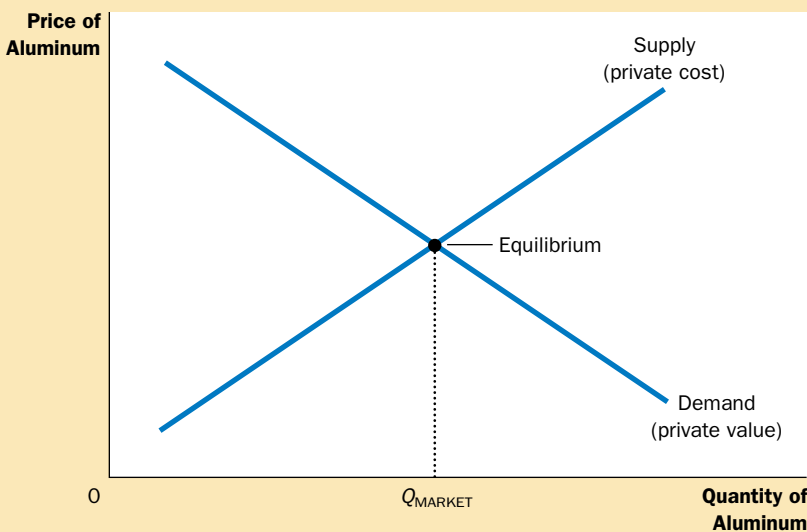


Figure 10-1

THE MARKET FOR ALUMINUM. The demand curve reflects the value to buyers, and the supply curve reflects the costs of sellers. The equilibrium quantity, Q_{MARKET} , maximizes the total value to buyers minus the total costs of sellers. In the absence of externalities, therefore, the market equilibrium is efficient.

market equilibrium, shown as Q_{MARKET} in Figure 10-1, is efficient in the sense that it maximizes the sum of producer and consumer surplus. That is, the market allocates resources in a way that maximizes the total value to the consumers who buy and use aluminum minus the total costs to the producers who make and sell aluminum.

NEGATIVE EXTERNALITIES IN PRODUCTION

Now let's suppose that aluminum factories emit pollution: For each unit of aluminum produced, a certain amount of smoke enters the atmosphere. Because this smoke creates a health risk for those who breathe the air, it is a negative externality. How does this externality affect the efficiency of the market outcome?

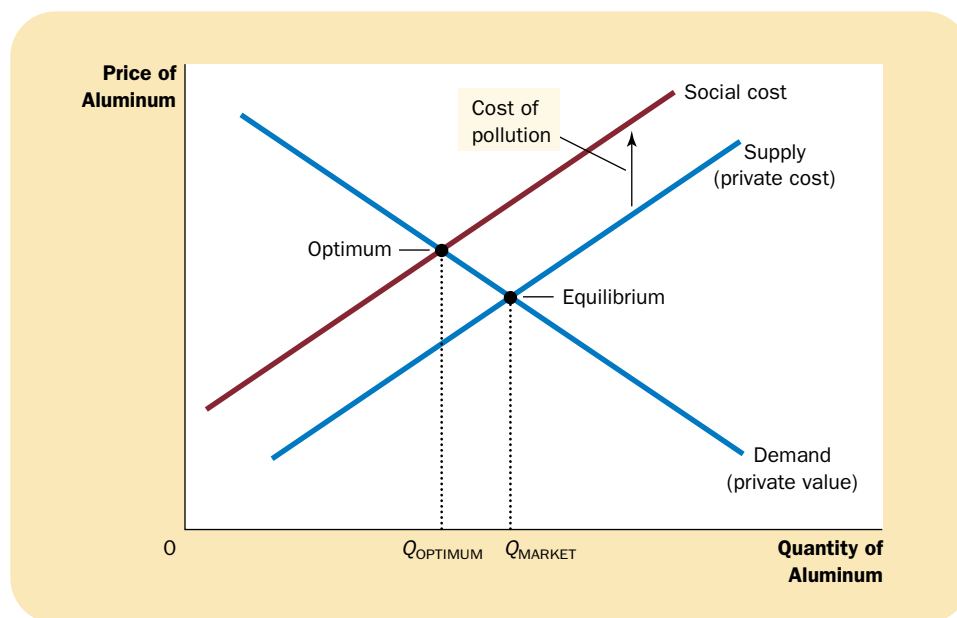
Because of the externality, the cost to *society* of producing aluminum is larger than the cost to the aluminum producers. For each unit of aluminum produced, the *social cost* includes the private costs of the aluminum producers plus the costs to those bystanders adversely affected by the pollution. Figure 10-2 shows the social cost of producing aluminum. The social-cost curve is above the supply curve because it takes into account the external costs imposed on society by aluminum producers. The difference between these two curves reflects the cost of the pollution emitted.

What quantity of aluminum should be produced? To answer this question, we once again consider what a benevolent social planner would do. The planner wants to maximize the total surplus derived from the market—the value to consumers of aluminum minus the cost of producing aluminum. The planner understands, however, that the cost of producing aluminum includes the external costs of the pollution.

The planner would choose the level of aluminum production at which the demand curve crosses the social-cost curve. This intersection determines the optimal amount of aluminum from the standpoint of society as a whole. Below this level of

Figure 10-2

POLLUTION AND THE SOCIAL OPTIMUM. In the presence of a negative externality to production, the social cost of producing aluminum exceeds the private cost. The optimal quantity of aluminum, Q_{OPTIMUM} , is therefore smaller than the equilibrium quantity, Q_{MARKET} .





"All I can say is that if being a leading manufacturer means being a leading polluter, so be it."

production, the value of the aluminum to consumers (as measured by the height of the demand curve) exceeds the social cost of producing it (as measured by the height of the social-cost curve). The planner does not produce more than this level because the social cost of producing additional aluminum exceeds the value to consumers.

Note that the equilibrium quantity of aluminum, Q_{MARKET} , is larger than the socially optimal quantity, Q_{OPTIMUM} . The reason for this inefficiency is that the market equilibrium reflects only the private costs of production. In the market equilibrium, the marginal consumer values aluminum at less than the social cost of producing it. That is, at Q_{MARKET} the demand curve lies below the social-cost curve. Thus, reducing aluminum production and consumption below the market equilibrium level raises total economic well-being.

How can the social planner achieve the optimal outcome? One way would be to tax aluminum producers for each ton of aluminum sold. The tax would shift the supply curve for aluminum upward by the size of the tax. If the tax accurately reflected the social cost of smoke released into the atmosphere, the new supply curve would coincide with the social-cost curve. In the new market equilibrium, aluminum producers would produce the socially optimal quantity of aluminum.

The use of such a tax is called **internalizing the externality** because it gives buyers and sellers in the market an incentive to take account of the external effects of their actions. Aluminum producers would, in essence, take the costs of pollution into account when deciding how much aluminum to supply because the tax now makes them pay for these external costs. Later in this chapter we consider other ways in which policymakers can deal with externalities.

internalizing an externality

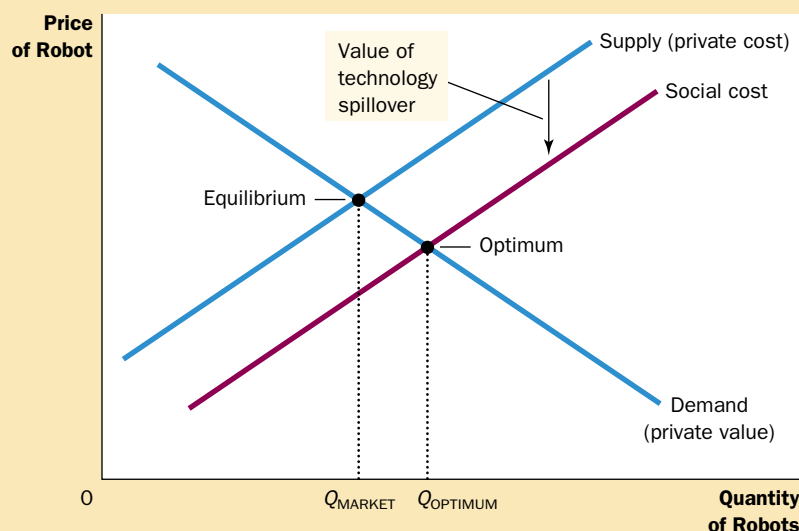
altering incentives so that people take account of the external effects of their actions

POSITIVE EXTERNALITIES IN PRODUCTION

Although in some markets the social cost of production exceeds the private cost, in other markets the opposite is the case. In these markets, the externality benefits bystanders, so the social cost of production is less than the private cost. One example is the market for industrial robots.

Figure 10-3

TECHNOLOGY SPILLOVERS AND THE SOCIAL OPTIMUM. In the presence of a positive externality to production, the social cost of producing robots is less than the private cost. The optimal quantity of robots, Q_{OPTIMUM} , is therefore larger than the equilibrium quantity, Q_{MARKET} .



Robots are at the frontier of a rapidly changing technology. Whenever a firm builds a robot, there is some chance that it will discover a new and better design. This new design will benefit not only this firm but society as a whole because the design will enter society's pool of technological knowledge. This type of positive externality is called a *technology spillover*.

The analysis of positive externalities is similar to the analysis of negative externalities. Figure 10-3 shows the market for robots. In this case, the social cost of production is less than the private cost reflected in the supply curve. In particular, the social cost of producing a robot is the private cost less the value of the technology spillover. Therefore, the social planner would choose to produce a larger quantity of robots than the private market does.

In this case, the government can internalize the externality by subsidizing the production of robots. If the government paid firms a subsidy for each robot produced, the supply curve would shift down by the amount of the subsidy, and this shift would increase the equilibrium quantity of robots. To ensure that the market equilibrium equals the social optimum, the subsidy should equal the value of the technology spillover.

CASE STUDY THE DEBATE OVER TECHNOLOGY POLICY

How large are technology spillovers, and what do they imply for public policy? This is an important question because technological progress is the key to why living standards rise from generation to generation. Yet it is also a difficult question on which economists often disagree.

Some economists believe that technology spillovers are pervasive and that the government should encourage those industries that yield the largest spillovers. For instance, these economists argue that if making computer chips

yields greater spillovers than making potato chips, then the government should use the tax laws to encourage the production of computer chips relative to the production of potato chips. Government intervention in the economy that aims to promote technology-enhancing industries is called *technology policy*.

Other economists are skeptical about technology policy. Even if technology spillovers are common, the success of a technology policy requires that the government be able to measure the size of the spillovers from different markets. This measurement problem is difficult at best. Moreover, without precise measurements, the political system may end up subsidizing those industries with the most political clout, rather than those that yield the largest positive externalities.

One type of technology policy that most economists endorse is patent protection. The patent laws protect the rights of inventors by giving them exclusive use of their inventions for a period of time. When a firm makes a technological breakthrough, it can patent the idea and capture much of the economic benefit for itself. The patent is said to internalize the externality by giving the firm a *property right* over its invention. If other firms want to use the new technology, they would have to obtain permission from the inventing firm and pay it some royalty. Thus, the patent system gives firms a greater incentive to engage in research and other activities that advance technology.

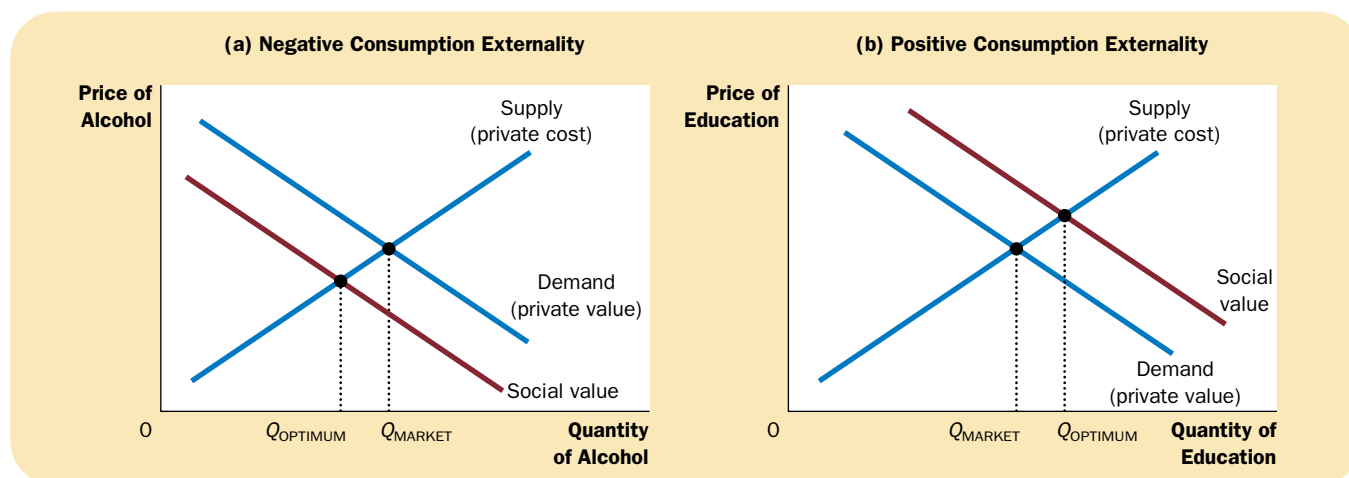
EXTERNALITIES IN CONSUMPTION

The externalities we have discussed so far are associated with the production of goods. Some externalities, however, are associated with consumption. The consumption of alcohol, for instance, yields negative externalities if consumers are more likely to drive under its influence and risk the lives of others. Similarly, the consumption of education yields positive externalities because a more educated population leads to better government, which benefits everyone.

The analysis of consumption externalities is similar to the analysis of production externalities. As Figure 10-4 shows, the demand curve does not reflect the value to society of the good. Panel (a) shows the case of a negative consumption externality, such as that associated with alcohol. In this case, the social value is less than the private value, and the socially optimal quantity is smaller than the quantity determined by the private market. Panel (b) shows the case of a positive consumption externality, like that of education. In this case, the social value is greater than the private value, and the socially optimal quantity is greater than the quantity determined by the private market.

Once again, the government can correct the market failure by inducing market participants to internalize the externality. The appropriate response in the case of consumption externalities is similar to that in the case of production externalities. To move the market equilibrium closer to the social optimum, a negative externality requires a tax, and a positive externality requires a subsidy. In fact, that is exactly the policy the government follows: Alcoholic beverages are among the most highly taxed goods in our economy, and education is heavily subsidized through public schools and government scholarships.

As you may have noticed, these examples of externalities lead to some general lessons: *Negative externalities in production or consumption lead markets to produce a larger quantity than is socially desirable. Positive externalities in production*

**Figure 10-4**

CONSUMPTION EXTERNALITIES. Panel (a) shows a market with a negative consumption externality, such as the market for alcoholic beverages. The curve representing social value is lower than the demand curve, and the socially optimal quantity, Q_{OPTIMUM} , is less than the equilibrium quantity, Q_{MARKET} . Panel (b) shows a market with a positive consumption externality, such as the market for education. The curve representing social value is above the demand curve, and the socially optimal quantity, Q_{OPTIMUM} , is greater than the equilibrium quantity, Q_{MARKET} .

or consumption lead markets to produce a smaller quantity than is socially desirable. To remedy the problem, the government can internalize the externality by taxing goods that have negative externalities and subsidizing goods that have positive externalities.

QUICK QUIZ: Give an example of a negative externality and a positive externality. ♦ Explain why market outcomes are inefficient in the presence of externalities.

PRIVATE SOLUTIONS TO EXTERNALITIES

We have discussed why externalities lead markets to allocate resources inefficiently, but have mentioned only briefly how this inefficiency can be remedied. In practice, both private actors and public policymakers respond to externalities in various ways. All of the remedies share the goal of moving the allocation of resources closer to the social optimum. In this section we examine private solutions.

THE TYPES OF PRIVATE SOLUTIONS

Although externalities tend to cause markets to be inefficient, government action is not always needed to solve the problem. In some circumstances, people can develop private solutions.

Sometimes, the problem of externalities is solved with moral codes and social sanctions. Consider, for instance, why most people do not litter. Although there are laws against littering, these laws are not vigorously enforced. Most people do not litter just because it is the wrong thing to do. The Golden Rule taught to most children says, “Do unto others as you would have them do unto you.” This moral injunction tells us to take account of how our actions affect other people. In economic terms, it tells us to internalize externalities.

Another private solution to externalities is charities, many of which are established to deal with externalities. For example, the Sierra Club, whose goal is to protect the environment, is a nonprofit organization funded with private donations. As another example, colleges and universities receive gifts from alumni, corporations, and foundations in part because education has positive externalities for society.

The private market can often solve the problem of externalities by relying on the self-interest of the relevant parties. Sometimes the solution takes the form of integrating different types of business. For example, consider an apple grower and a beekeeper that are located next to each other. Each business confers a positive externality on the other: By pollinating the flowers on the trees, the bees help the orchard produce apples. At the same time, the bees use the nectar they get from the apple trees to produce honey. Nonetheless, when the apple grower is deciding how many trees to plant and the beekeeper is deciding how many bees to keep, they neglect the positive externality. As a result, the apple grower plants too few trees and the beekeeper keeps too few bees. These externalities could be internalized if the beekeeper bought the apple orchard or if the apple grower bought the beehive: Both activities would then take place within the same firm, and this single firm could choose the optimal number of trees and bees. Internalizing externalities is one reason that some firms are involved in different types of business.

Another way for the private market to deal with external effects is for the interested parties to enter into a contract. In the foregoing example, a contract between the apple grower and the beekeeper can solve the problem of too few trees and too few bees. The contract can specify the number of trees, the number of bees, and perhaps a payment from one party to the other. By setting the right number of trees and bees, the contract can solve the inefficiency that normally arises from these externalities and make both parties better off.

THE COASE THEOREM

How effective is the private market in dealing with externalities? A famous result, called the **Coase theorem** after economist Ronald Coase, suggests that it can be very effective in some circumstances. According to the Coase theorem, if private parties can bargain without cost over the allocation of resources, then the private market will always solve the problem of externalities and allocate resources efficiently.

To see how the Coase theorem works, consider an example. Suppose that Dick owns a dog named Spot. Spot barks and disturbs Jane, Dick’s neighbor. Dick gets a benefit from owning the dog, but the dog confers a negative externality on Jane. Should Dick be forced to send Spot to the pound, or should Jane have to suffer sleepless nights because of Spot’s barking?

Consider first what outcome is socially efficient. A social planner, considering the two alternatives, would compare the benefit that Dick gets from the dog to the cost that Jane bears from the barking. If the benefit exceeds the cost, it is efficient

Coase theorem

the proposition that if private parties can bargain without cost over the allocation of resources, they can solve the problem of externalities on their own

for Dick to keep the dog and for Jane to live with the barking. Yet if the cost exceeds the benefit, then Dick should get rid of the dog.

According to the Coase theorem, the private market will reach the efficient outcome on its own. How? Jane can simply offer to pay Dick to get rid of the dog. Dick will accept the deal if the amount of money Jane offers is greater than the benefit of keeping the dog.

By bargaining over the price, Dick and Jane can always reach the efficient outcome. For instance, suppose that Dick gets a \$500 benefit from the dog and Jane bears an \$800 cost from the barking. In this case, Jane can offer Dick \$600 to get rid of the dog, and Dick will gladly accept. Both parties are better off than they were before, and the efficient outcome is reached.

It is possible, of course, that Jane would not be willing to offer any price that Dick would accept. For instance, suppose that Dick gets a \$1,000 benefit from the dog and Jane bears an \$800 cost from the barking. In this case, Dick would turn down any offer below \$1,000, while Jane would not offer any amount above \$800. Therefore, Dick ends up keeping the dog. Given these costs and benefits, however, this outcome is efficient.

So far, we have assumed that Dick has the legal right to keep a barking dog. In other words, we have assumed that Dick can keep Spot unless Jane pays him enough to induce him to give up the dog voluntarily. How different would the outcome be, on the other hand, if Jane had the legal right to peace and quiet?

According to the Coase theorem, the initial distribution of rights does not matter for the market's ability to reach the efficient outcome. For instance, suppose that Jane can legally compel Dick to get rid of the dog. Although having this right works to Jane's advantage, it probably will not change the outcome. In this case, Dick can offer to pay Jane to allow him to keep the dog. If the benefit of the dog to Dick exceeds the cost of the barking to Jane, then Dick and Jane will strike a bargain in which Dick keeps the dog.

Although Dick and Jane can reach the efficient outcome regardless of how rights are initially distributed, the distribution of rights is not irrelevant: It determines the distribution of economic well-being. Whether Dick has the right to a barking dog or Jane the right to peace and quiet determines who pays whom in the final bargain. But, in either case, the two parties can bargain with each other and solve the externality problem. Dick will end up keeping the dog only if the benefit exceeds the cost.

To sum up: *The Coase theorem says that private economic actors can solve the problem of externalities among themselves. Whatever the initial distribution of rights, the interested parties can always reach a bargain in which everyone is better off and the outcome is efficient.*

WHY PRIVATE SOLUTIONS DO NOT ALWAYS WORK

Despite the appealing logic of the Coase theorem, private actors on their own often fail to resolve the problems caused by externalities. The Coase theorem applies only when the interested parties have no trouble reaching and enforcing an agreement. In the real world, however, bargaining does not always work, even when a mutually beneficial agreement is possible.

Sometimes the interested parties fail to solve an externality problem because of **transaction costs**, the costs that parties incur in the process of agreeing to and following through on a bargain. In our example, imagine that Dick and Jane speak

transaction costs

the costs that parties incur in the process of agreeing and following through on a bargain

different languages so that, to reach an agreement, they will need to hire a translator. If the benefit of solving the barking problem is less than the cost of the translator, Dick and Jane might choose to leave the problem unsolved. In more realistic examples, the transaction costs are the expenses not of translators but of the lawyers required to draft and enforce contracts.

Other times bargaining simply breaks down. The recurrence of wars and labor strikes shows that reaching agreement can be difficult and that failing to reach agreement can be costly. The problem is often that each party tries to hold out for a better deal. For example, suppose that Dick gets a \$500 benefit from the dog, and Jane bears an \$800 cost from the barking. Although it is efficient for Jane to pay Dick to get rid of the dog, there are many prices that could lead to this outcome. Dick might demand \$750, and Jane might offer only \$550. As they haggle over the price, the inefficient outcome with the barking dog persists.

Reaching an efficient bargain is especially difficult when the number of interested parties is large because coordinating everyone is costly. For example, consider a factory that pollutes the water of a nearby lake. The pollution confers a negative externality on the local fishermen. According to the Coase theorem, if the pollution is inefficient, then the factory and the fishermen could reach a bargain in which the fishermen pay the factory not to pollute. If there are many fishermen, however, trying to coordinate them all to bargain with the factory may be almost impossible.

When private bargaining does not work, the government can sometimes play a role. The government is an institution designed for collective action. In this example, the government can act on behalf of the fishermen, even when it is impractical for the fishermen to act for themselves. In the next section, we examine how the government can try to remedy the problem of externalities.

QUICK QUIZ: Give an example of a private solution to an externality.

- ◆ What is the Coase theorem? ◆ Why are private economic actors sometimes unable to solve the problems caused by an externality?

PUBLIC POLICIES TOWARD EXTERNALITIES

When an externality causes a market to reach an inefficient allocation of resources, the government can respond in one of two ways. *Command-and-control policies* regulate behavior directly. *Market-based policies* provide incentives so that private decisionmakers will choose to solve the problem on their own.

REGULATION

The government can remedy an externality by making certain behaviors either required or forbidden. For example, it is a crime to dump poisonous chemicals into the water supply. In this case, the external costs to society far exceed the benefits to the polluter. The government therefore institutes a command-and-control policy that prohibits this act altogether.

In most cases of pollution, however, the situation is not this simple. Despite the stated goals of some environmentalists, it would be impossible to prohibit all

polluting activity. For example, virtually all forms of transportation—even the horse—produce some undesirable polluting by-products. But it would not be sensible for the government to ban all transportation. Thus, instead of trying to eradicate pollution altogether, society has to weigh the costs and benefits to decide the kinds and quantities of pollution it will allow. In the United States, the Environmental Protection Agency (EPA) is the government agency with the task of developing and enforcing regulations aimed at protecting the environment.

Environmental regulations can take many forms. Sometimes the EPA dictates a maximum level of pollution that a factory may emit. Other times the EPA requires that firms adopt a particular technology to reduce emissions. In all cases, to design good rules, the government regulators need to know the details about specific industries and about the alternative technologies that those industries could adopt. This information is often difficult for government regulators to obtain.

PIGOVIAN TAXES AND SUBSIDIES

Instead of regulating behavior in response to an externality, the government can use market-based policies to align private incentives with social efficiency. For instance, as we saw earlier, the government can internalize the externality by taxing activities that have negative externalities and subsidizing activities that have positive externalities. Taxes enacted to correct the effects of negative externalities are called **Pigovian taxes**, after economist Arthur Pigou (1877–1959), an early advocate of their use.

Economists usually prefer Pigovian taxes over regulations as a way to deal with pollution because they can reduce pollution at a lower cost to society. To see why, let us consider an example.

Suppose that two factories—a paper mill and a steel mill—are each dumping 500 tons of glop into a river each year. The EPA decides that it wants to reduce the amount of pollution. It considers two solutions:

- ◆ *Regulation:* The EPA could tell each factory to reduce its pollution to 300 tons of glop per year.
- ◆ *Pigovian tax:* The EPA could levy a tax on each factory of \$50,000 for each ton of glop it emits.

The regulation would dictate a level of pollution, whereas the tax would give factory owners an economic incentive to reduce pollution. Which solution do you think is better?

Most economists would prefer the tax. They would first point out that a tax is just as effective as a regulation in reducing the overall level of pollution. The EPA can achieve whatever level of pollution it wants by setting the tax at the appropriate level. The higher the tax, the larger the reduction in pollution. Indeed, if the tax is high enough, the factories will close down altogether, reducing pollution to zero.

The reason why economists would prefer the tax is that it reduces pollution more efficiently. The regulation requires each factory to reduce pollution by the same amount, but an equal reduction is not necessarily the least expensive way to clean up the water. It is possible that the paper mill can reduce pollution at lower cost than the steel mill. If so, the paper mill would respond to the tax by reducing pollution substantially to avoid the tax, whereas the steel mill would respond by reducing pollution less and paying the tax.

Pigovian tax

a tax enacted to correct the effects of a negative externality

In essence, the Pigovian tax places a price on the right to pollute. Just as markets allocate goods to those buyers who value them most highly, a Pigovian tax allocates pollution to those factories that face the highest cost of reducing it. Whatever the level of pollution the EPA chooses, it can achieve this goal at the lowest total cost using a tax.

Economists also argue that Pigovian taxes are better for the environment. Under the command-and-control policy of regulation, factories have no reason to reduce emission further once they have reached the target of 300 tons of glop. By contrast, the tax gives the factories an incentive to develop cleaner technologies, because a cleaner technology would reduce the amount of tax the factory has to pay.

Pigovian taxes are unlike most other taxes. As we discussed in Chapter 8, most taxes distort incentives and move the allocation of resources away from the social optimum. The reduction in economic well-being—that is, in consumer and producer surplus—exceeds the amount of revenue the government raises, resulting in a deadweight loss. By contrast, when externalities are present, society also cares about the well-being of the bystanders who are affected. Pigovian taxes correct incentives for the presence of externalities and thereby move the allocation of resources closer to the social optimum. Thus, while Pigovian taxes raise revenue for the government, they enhance economic efficiency.

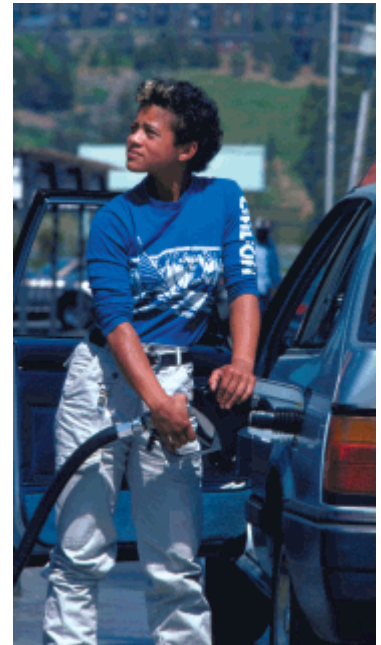
CASE STUDY WHY IS GASOLINE TAXED SO HEAVILY?

In many countries, gasoline is among the most heavily taxed goods in the economy. In the United States, for instance, almost half of what drivers pay for gasoline goes to the gas tax. In many European countries, the tax is even larger and the price of gasoline is three or four times the U.S. price.

Why is this tax so common? One answer is that the gas tax is a Pigovian tax aimed at correcting three negative externalities associated with driving:

- ◆ *Congestion:* If you have ever been stuck in bumper-to-bumper traffic, you have probably wished that there were fewer cars on the road. A gasoline tax keeps congestion down by encouraging people to take public transportation, car pool more often, and live closer to work.
- ◆ *Accidents:* Whenever a person buys a large car or a sport utility vehicle, he makes himself safer, but he puts his neighbors at risk. According to the National Highway Traffic Safety Administration, a person driving a typical car is five times as likely to die if hit by a sport utility vehicle than if hit by another car. The gas tax is an indirect way of making people pay when their large, gas-guzzling vehicles impose risk on others, which in turn makes them take account of this risk when choosing what vehicle to purchase.
- ◆ *Pollution:* The burning of fossil fuels such as gasoline is widely believed to be the cause of global warming. Experts disagree about how dangerous this threat is, but there is no doubt that the gas tax reduces the risk by reducing the use of gasoline.

So the gas tax, rather than causing deadweight losses like most taxes, actually makes the economy work better. It means less traffic congestion, safer roads, and a cleaner environment.



"If the gas tax were any larger, I'd take the bus."

TRADABLE POLLUTION PERMITS

Returning to our example of the paper mill and the steel mill, let us suppose that, despite the advice of its economists, the EPA adopts the regulation and requires each factory to reduce its pollution to 300 tons of glop per year. Then one day, after the regulation is in place and both mills have complied, the two firms go to the EPA with a proposal. The steel mill wants to increase its emission of glop by 100 tons. The paper mill has agreed to reduce its emission by the same amount if the steel mill pays it \$5 million. Should the EPA allow the two factories to make this deal?

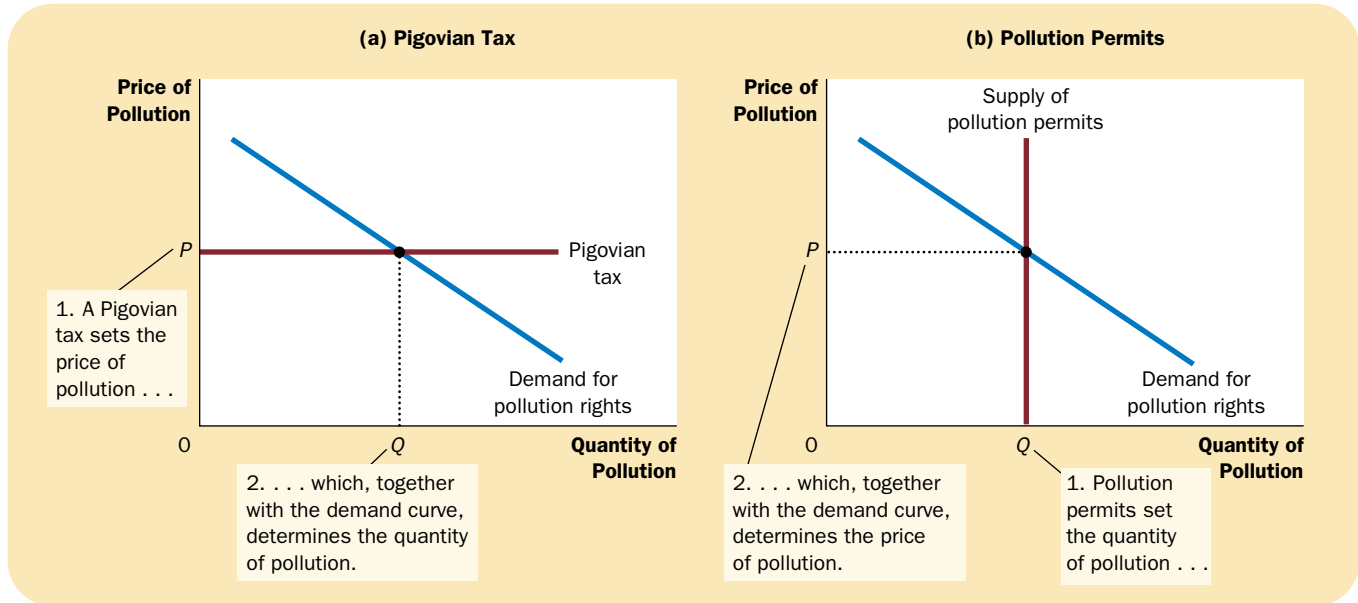
From the standpoint of economic efficiency, allowing the deal is good policy. The deal must make the owners of the two factories better off, because they are voluntarily agreeing to it. Moreover, the deal does not have any external effects because the total amount of pollution remains the same. Thus, social welfare is enhanced by allowing the paper mill to sell its right to pollute to the steel mill.

The same logic applies to any voluntary transfer of the right to pollute from one firm to another. If the EPA allows firms to make these deals, it will, in essence, have created a new scarce resource: pollution permits. A market to trade these permits will eventually develop, and that market will be governed by the forces of supply and demand. The invisible hand will ensure that this new market efficiently allocates the right to pollute. The firms that can reduce pollution only at high cost will be willing to pay the most for the pollution permits. The firms that can reduce pollution at low cost will prefer to sell whatever permits they have.

One advantage of allowing a market for pollution permits is that the initial allocation of pollution permits among firms does not matter from the standpoint of economic efficiency. The logic behind this conclusion is similar to that behind the Coase theorem. Those firms that can reduce pollution most easily would be willing to sell whatever permits they get, and those firms that can reduce pollution only at high cost would be willing to buy whatever permits they need. As long as there is a free market for the pollution rights, the final allocation will be efficient whatever the initial allocation.

Although reducing pollution using pollution permits may seem quite different from using Pigovian taxes, in fact the two policies have much in common. In both cases, firms pay for their pollution. With Pigovian taxes, polluting firms must pay a tax to the government. With pollution permits, polluting firms must pay to buy the permit. (Even firms that already own permits must pay to pollute: The opportunity cost of polluting is what they could have received by selling their permits on the open market.) Both Pigovian taxes and pollution permits internalize the externality of pollution by making it costly for firms to pollute.

The similarity of the two policies can be seen by considering the market for pollution. Both panels in Figure 10-5 show the demand curve for the right to pollute. This curve shows that the lower the price of polluting, the more firms will choose to pollute. In panel (a), the EPA uses a Pigovian tax to set a price for pollution. In this case, the supply curve for pollution rights is perfectly elastic (because firms can pollute as much as they want by paying the tax), and the position of the demand curve determines the quantity of pollution. In panel (b), the EPA sets a quantity of pollution by issuing pollution permits. In this case, the supply curve for pollution rights is perfectly inelastic (because the quantity of pollution is fixed by the number of permits), and the position of the demand curve determines the price of pollution. Hence, for any given demand curve for pollution, the EPA can



THE EQUIVALENCE OF PIGOVIAN TAXES AND POLLUTION PERMITS. In panel (a), the EPA sets a price on pollution by levying a Pigovian tax, and the demand curve determines the quantity of pollution. In panel (b), the EPA limits the quantity of pollution by limiting the number of pollution permits, and the demand curve determines the price of pollution. The price and quantity of pollution are the same in the two cases.

Figure 10-5

achieve any point on the demand curve either by setting a price with a Pigovian tax or by setting a quantity with pollution permits.

In some circumstances, however, selling pollution permits may be better than levying a Pigovian tax. Suppose the EPA wants no more than 600 tons of glop to be dumped into the river. But, because the EPA does not know the demand curve for pollution, it is not sure what size tax would achieve that goal. In this case, it can simply auction off 600 pollution permits. The auction price would yield the appropriate size of the Pigovian tax.

The idea of the government auctioning off the right to pollute may at first sound like a creature of some economist's imagination. And, in fact, that is how the idea began. But increasingly the EPA has used the system as a way to control pollution. Pollution permits, like Pigovian taxes, are now widely viewed as a cost-effective way to keep the environment clean.

OBJECTIONS TO THE ECONOMIC ANALYSIS OF POLLUTION

"We cannot give anyone the option of polluting for a fee." This comment by former Senator Edmund Muskie reflects the view of some environmentalists. Clean air and clean water, they argue, are fundamental human rights that should not be debased by considering them in economic terms. How can you put a price on clean air and clean water? The environment is so important, they claim, that we should protect it as much as possible, regardless of the cost.



Economists have little sympathy with this type of argument. To economists, good environmental policy begins by acknowledging the first of the *Ten Principles of Economics* in Chapter 1: People face tradeoffs. Certainly, clean air and clean water have value. But their value must be compared to their opportunity cost—that is, to what one must give up to obtain them. Eliminating all pollution is impossible. Trying to eliminate all pollution would reverse many of the technological advances that allow us to enjoy a high standard of living. Few people would be willing to accept poor nutrition, inadequate medical care, or shoddy housing to make the environment as clean as possible.

Economists argue that some environmental activists hurt their own cause by not thinking in economic terms. A clean environment is a good like other goods. Like all normal goods, it has a positive income elasticity: Rich countries can afford a cleaner environment than poor ones and, therefore, usually have more rigorous environmental protection. In addition, like most other goods, clean air and water obey the law of demand: The lower the price of environmental protection, the more the public will want. The economic approach of using pollution permits and Pigovian taxes reduces the cost of environmental protection and should, therefore, increase the public's demand for a clean environment.

QUICK QUIZ: A glue factory and a steel mill emit smoke containing a chemical that is harmful if inhaled in large amounts. Describe three ways the town government might respond to this externality. What are the pros and cons of each of your solutions?

CONCLUSION

The invisible hand is powerful but not omnipotent. A market's equilibrium maximizes the sum of producer and consumer surplus. When the buyers and sellers in the market are the only interested parties, this outcome is efficient from the standpoint of society as a whole. But when there are external effects, such as pollution, evaluating a market outcome requires taking into account the well-being of third parties as well. In this case, the invisible hand of the marketplace may fail to allocate resources efficiently.

In some cases, people can solve the problem of externalities on their own. The Coase theorem suggests that the interested parties can bargain among themselves and agree on an efficient solution. Sometimes, however, an efficient outcome cannot be reached, perhaps because the large number of interested parties makes bargaining difficult.

When people cannot solve the problem of externalities privately, the government often steps in. Yet, even now, society should not abandon market forces entirely. Rather, the government can address the problem by requiring decision-makers to bear the full costs of their actions. Pigovian taxes on emissions and pollution permits, for instance, are designed to internalize the externality of pollution. More and more, they are the policy of choice for those interested in protecting the environment. Market forces, properly redirected, are often the best remedy for market failure.

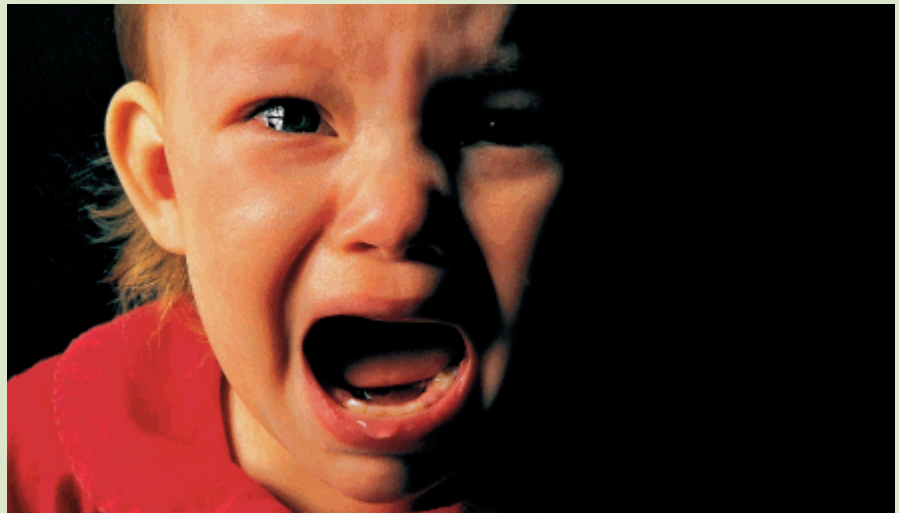
IN THE NEWS*Children as Externalities*

THIS TONGUE-IN-CHEEK EDITORIAL FROM *THE Economist*, an international news-magazine, calls attention to a common externality that is not fully appreciated.

**Mum's the Word:
When Children Should Be
Screened and Not Heard**

We live in increasingly intolerant times. Signs proliferate demanding no smoking, no spitting, no parking, even no walking. . . . Posh clubs and restaurants have long had "no jeans" rules, but these days you can be too smart. Some London hostels have "no suits" policies, for fear that boisterous city traders in suits might spoil the atmosphere. Environmentalists have long demanded all sorts of bans on cars. Mobile telephones are the latest target: some trains, airline lounges, restaurants, and even golf courses are being designated "no phone" areas.

If intolerance really has to be the spirit of this age, *The Economist* would like to suggest restrictions on another source of noise pollution: children. Lest you dismiss this as mere prejudice, we can even produce a good economic argument for it. Smoking, driving, and mobile phones all cause what economists call "negative externalities." That is, the costs of these activities to other people tend to exceed the costs to the individuals



A NEGATIVE EXTERNALITY

of their proclivities. The invisible hand of the market fumbles, leading resources astray. Thus, because a driver's private motoring costs do not reflect the costs he imposes on others in the form of pollution and congestion, he uses the car more than is socially desirable. Likewise, it is argued, smokers take too little care to ensure that their acrid fumes do not damage other people around them.

Governments typically respond to such market failures in two ways. One is higher taxes, to make polluters pay the full cost of their anti-social behavior. The other is regulation, such as emission standards or bans on smoking in public places. Both approaches might work for children.

For children, just like cigarettes or mobile phones, clearly impose a negative externality on people who are near them. Anybody who has suffered a 12-hour flight with a bawling baby in the row immediately ahead, or a bored youngster viciously kicking their seat from behind, will grasp this as quickly as they would love to grasp the youngster's neck. Here is a clear case of market failure: parents

do not bear the full costs (indeed young babies travel free), so they are too ready to take their noisy brats with them. Where is the invisible hand when it is needed to administer a good smack?

The solution is obvious. All airlines, trains, and restaurants should create child-free zones. Put all those children at the back of the plane and parents might make more effort to minimize their noise pollution. And instead of letting children pay less and babies go free, they should be charged (or taxed) more than adults, with the revenues used to subsidize seats immediately in front of the war-zone.

Passengers could then request a no-children seat, just as they now ask for a no-smoking one. As more women choose not to have children and the number of older people without young children increases, the demand for child-free travel will expand. Well, yes, it is a bit intolerant—but why shouldn't parents be treated as badly as smokers? And at least there is an obvious airline to pioneer the scheme: Virgin.

SOURCE: *The Economist*, December 5, 1998, p. 20.

Summary

- ◆ When a transaction between a buyer and seller directly affects a third party, the effect is called an externality. Negative externalities, such as pollution, cause the socially optimal quantity in a market to be less than the equilibrium quantity. Positive externalities, such as technology spillovers, cause the socially optimal quantity to be greater than the equilibrium quantity.
- ◆ Those affected by externalities can sometimes solve the problem privately. For instance, when one business confers an externality on another business, the two businesses can internalize the externality by merging. Alternatively, the interested parties can solve the problem by negotiating a contract. According to the Coase theorem, if people can bargain without cost, then they can always reach an agreement in which resources are allocated efficiently. In many cases, however, reaching a bargain among the many interested parties is difficult, so the Coase theorem does not apply.
- ◆ When private parties cannot adequately deal with external effects, such as pollution, the government often steps in. Sometimes the government prevents socially inefficient activity by regulating behavior. Other times it internalizes an externality using Pigovian taxes. Another way to protect the environment is for the government to issue a limited number of pollution permits. The end result of this policy is largely the same as imposing Pigovian taxes on polluters.

Key Concepts

externality, p. 206

internalizing an externality, p. 209

Coase theorem, p. 213

transaction costs, p. 214

Pigovian tax, p. 216

Questions for Review

1. Give an example of a negative externality and an example of a positive externality.
2. Use a supply-and-demand diagram to explain the effect of a negative externality in production.
3. In what way does the patent system help society solve an externality problem?
4. List some of the ways that the problems caused by externalities can be solved without government intervention.
5. Imagine that you are a nonsmoker sharing a room with a smoker. According to the Coase theorem, what determines whether your roommate smokes in the room? Is this outcome efficient? How do you and your roommate reach this solution?
6. What are Pigovian taxes? Why do economists prefer them over regulations as a way to protect the environment from pollution?

Problems and Applications

1. Do you agree with the following statements? Why or why not?
 - a. "The benefits of Pigovian taxes as a way to reduce pollution have to be weighed against the deadweight losses that these taxes cause."
 - b. "A negative production externality calls for a Pigovian tax on producers, whereas a negative consumption externality calls for a Pigovian tax on consumers."
2. Consider the market for fire extinguishers.
 - a. Why might fire extinguishers exhibit positive externalities in consumption?
 - b. Draw a graph of the market for fire extinguishers, labeling the demand curve, the social-value

- curve, the supply curve, and the social-cost curve.
- c. Indicate the market equilibrium level of output and the efficient level of output. Give an intuitive explanation for why these quantities differ.
 - d. If the external benefit is \$10 per extinguisher, describe a government policy that would result in the efficient outcome.
3. Contributions to charitable organizations are deductible under the federal income tax. In what way does this government policy encourage private solutions to externalities?
 4. Ringo loves playing rock and roll music at high volume. Luciano loves opera and hates rock and roll. Unfortunately, they are next-door neighbors in an apartment building with paper-thin walls.
 - a. What is the externality here?
 - b. What command-and-control policy might the landlord impose? Could such a policy lead to an inefficient outcome?
 - c. Suppose the landlord lets the tenants do whatever they want. According to the Coase theorem, how might Ringo and Luciano reach an efficient outcome on their own? What might prevent them from reaching an efficient outcome?
 5. It is rumored that the Swiss government subsidizes cattle farming, and that the subsidy is larger in areas with more tourist attractions. Can you think of a reason why this policy might be efficient?
 6. Greater consumption of alcohol leads to more motor vehicle accidents and, thus, imposes costs on people who do not drink and drive.
 - a. Illustrate the market for alcohol, labeling the demand curve, the social-value curve, the supply curve, the social-cost curve, the market equilibrium level of output, and the efficient level of output.
 - b. On your graph, shade the area corresponding to the deadweight loss of the market equilibrium. (Hint: The deadweight loss occurs because some units of alcohol are consumed for which the social cost exceeds the social value.) Explain.
 7. Many observers believe that the levels of pollution in our economy are too high.
 - a. If society wishes to reduce overall pollution by a certain amount, why is it efficient to have different amounts of reduction at different firms?
 - b. Command-and-control approaches often rely on uniform reductions among firms. Why are these approaches generally unable to target the firms that should undertake bigger reductions?
 - c. Economists argue that appropriate Pigovian taxes or tradable pollution rights will result in efficient pollution reduction. How do these approaches target the firms that should undertake bigger reductions?
 8. The Pristine River has two polluting firms on its banks. Acme Industrial and Creative Chemicals each dump 100 tons of glop into the river each year. The cost of reducing glop emissions per ton equals \$10 for Acme and \$100 for Creative. The local government wants to reduce overall pollution from 200 tons to 50 tons.
 - a. If the government knew the cost of reduction for each firm, what reductions would it impose to reach its overall goal? What would be the cost to each firm and the total cost to the firms together?
 - b. In a more typical situation, the government would not know the cost of pollution reduction at each firm. If the government decided to reach its overall goal by imposing uniform reductions on the firms, calculate the reduction made by each firm, the cost to each firm, and the total cost to the firms together.
 - c. Compare the total cost of pollution reduction in parts (a) and (b). If the government does not know the cost of reduction for each firm, is there still some way for it to reduce pollution to 50 tons at the total cost you calculated in part (a)? Explain.
 9. Figure 10-5 shows that for any given demand curve for the right to pollute, the government can achieve the same outcome either by setting a price with a Pigovian tax or by setting a quantity with pollution permits. Suppose there is a sharp improvement in the technology for controlling pollution.
 - a. Using graphs similar to those in Figure 10-5, illustrate the effect of this development on the demand for pollution rights.
 - b. What is the effect on the price and quantity of pollution under each regulatory system? Explain.
 10. Suppose that the government decides to issue tradable permits for a certain form of pollution.
 - a. Does it matter for economic efficiency whether the government distributes or auctions the permits? Does it matter in any other ways?
 - b. If the government chooses to distribute the permits, does the allocation of permits among firms matter for efficiency? Does it matter in any other ways?
 11. The primary cause of global warming is carbon dioxide, which enters the atmosphere in varying amounts from

different countries but is distributed equally around the globe within a year. In an article in *The Boston Globe* (July 3, 1990), Martin and Kathleen Feldstein argue that the correct approach to global warming is “not to ask individual countries to stabilize their emissions of carbon dioxide at current levels,” as some have suggested. Instead, they argue that “carbon dioxide emissions should be reduced in countries where the costs are least, and the countries that bear that burden should be compensated by the rest of the world.”

- a. Why is international cooperation necessary to reach an efficient outcome?
 - b. Is it possible to devise a compensation scheme such that all countries would be better off than under a system of uniform emission reductions? Explain.
12. Some people object to market-based policies to reduce pollution, claiming that they place a dollar value on cleaning our air and water. Economists reply that society *implicitly* places a dollar value on environmental cleanup even under command-and-control policies. Discuss why this is true.

13. (This problem is challenging.) There are three industrial firms in Happy Valley.

FIRM	INITIAL POLLUTION LEVEL	COST OF REDUCING POLLUTION BY 1 UNIT
A	70 units	\$20
B	80	25
C	50	10

The government wants to reduce pollution to 120 units, so it gives each firm 40 tradable pollution permits.

- a. Who sells permits and how many do they sell? Who buys permits and how many do they buy? Briefly explain why the sellers and buyers are each willing to do so. What is the total cost of pollution reduction in this situation?
- b. How much higher would the costs of pollution reduction be if the permits could not be traded?

11



PUBLIC GOODS AND COMMON RESOURCES

An old song lyric maintains that “the best things in life are free.” A moment’s thought reveals a long list of goods that the songwriter could have had in mind. Nature provides some of them, such as rivers, mountains, beaches, lakes, and oceans. The government provides others, such as playgrounds, parks, and parades. In each case, people do not pay a fee when they choose to enjoy the benefit of the good.

Free goods provide a special challenge for economic analysis. Most goods in our economy are allocated in markets, where buyers pay for what they receive and sellers are paid for what they provide. For these goods, prices are the signals that guide the decisions of buyers and sellers. When goods are available free of charge, however, the market forces that normally allocate resources in our economy are absent.

In this chapter we examine the problems that arise for goods without market prices. Our analysis will shed light on one of the *Ten Principles of Economics*

IN THIS CHAPTER YOU WILL . . .

Learn the defining characteristics of public goods and common resources

Examine why private markets fail to provide public goods

Consider some of the important public goods in our economy

See why the cost-benefit analysis of public goods is both necessary and difficult

Examine why people tend to use common resources too much

Consider some of the important common resources in our economy



in Chapter 1: Governments can sometimes improve market outcomes. When a good does not have a price attached to it, private markets cannot ensure that the good is produced and consumed in the proper amounts. In such cases, government policy can potentially remedy the market failure and raise economic well-being.

THE DIFFERENT KINDS OF GOODS

How well do markets work in providing the goods that people want? The answer to this question depends on the good being considered. As we discussed in Chapter 7, we can rely on the market to provide the efficient number of ice-cream cones: The price of ice-cream cones adjusts to balance supply and demand, and this equilibrium maximizes the sum of producer and consumer surplus. Yet, as we discussed in Chapter 10, we cannot rely on the market to prevent aluminum manufacturers from polluting the air we breathe: Buyers and sellers in a market typically do not take account of the external effects of their decisions. Thus, markets work well when the good is ice cream, but they work badly when the good is clean air.

In thinking about the various goods in the economy, it is useful to group them according to two characteristics:

excludability

the property of a good whereby a person can be prevented from using it

rivalry

the property of a good whereby one person's use diminishes other people's use

private goods

goods that are both excludable and rival

public goods

goods that are neither excludable nor rival

common resources

goods that are rival but not excludable

- ◆ Is the good **excludable**? Can people be prevented from using the good?
- ◆ Is the good **rival**? Does one person's use of the good diminish another person's enjoyment of it?

Using these two characteristics, Figure 11-1 divides goods into four categories:

1. **Private goods** are both excludable and rival. Consider an ice-cream cone, for example. An ice-cream cone is excludable because it is possible to prevent someone from eating an ice-cream cone—you just don't give it to him. An ice-cream cone is rival because if one person eats an ice-cream cone, another person cannot eat the same cone. Most goods in the economy are private goods like ice-cream cones. When we analyzed supply and demand in Chapters 4, 5, and 6 and the efficiency of markets in Chapters 7, 8, and 9, we implicitly assumed that goods were both excludable and rival.
2. **Public goods** are neither excludable nor rival. That is, people cannot be prevented from using a public good, and one person's enjoyment of a public good does not reduce another person's enjoyment of it. For example, national defense is a public good. Once the country is defended from foreign aggressors, it is impossible to prevent any single person from enjoying the benefit of this defense. Moreover, when one person enjoys the benefit of national defense, he does not reduce the benefit to anyone else.
3. **Common resources** are rival but not excludable. For example, fish in the ocean are a rival good: When one person catches fish, there are fewer fish for the next person to catch. Yet these fish are not an excludable good because it is difficult to charge fishermen for the fish that they catch.
4. When a good is excludable but not rival, it is an example of a *natural monopoly*. For instance, consider fire protection in a small town. It is easy to

Figure 11-1

		Rival?	
		Yes	No
Excludable?	Yes	Private Goods • Ice-cream cones • Clothing • Congested toll roads	Natural Monopolies • Fire protection • Cable TV • Uncongested toll roads
	No	Common Resources • Fish in the ocean • The environment • Congested nontoll roads	Public Goods • National defense • Knowledge • Uncongested nontoll roads

FOUR TYPES OF GOODS.

Goods can be grouped into four categories according to two questions: (1) Is the good excludable? That is, can people be prevented from using it? (2) Is the good rival? That is, does one person's use of the good diminish other people's use of it? This table gives examples of goods in each of the four categories.

exclude people from enjoying this good: The fire department can just let their house burn down. Yet fire protection is not rival. Firefighters spend much of their time waiting for a fire, so protecting an extra house is unlikely to reduce the protection available to others. In other words, once a town has paid for the fire department, the additional cost of protecting one more house is small. In Chapter 15 we give a more complete definition of natural monopolies and study them in some detail.

In this chapter we examine goods that are not excludable and, therefore, are available to everyone free of charge: public goods and common resources. As we will see, this topic is closely related to the study of externalities. For both public goods and common resources, externalities arise because something of value has no price attached to it. If one person were to provide a public good, such as national defense, other people would be better off, and yet they could not be charged for this benefit. Similarly, when one person uses a common resource, such as the fish in the ocean, other people are worse off, and yet they are not compensated for this loss. Because of these external effects, private decisions about consumption and production can lead to an inefficient allocation of resources, and government intervention can potentially raise economic well-being.

QUICK QUIZ: Define *public goods* and *common resources*, and give an example of each.

PUBLIC GOODS

To understand how public goods differ from other goods and what problems they present for society, let's consider an example: a fireworks display. This good is not excludable because it is impossible to prevent someone from seeing fireworks, and it is not rival because one person's enjoyment of fireworks does not reduce anyone else's enjoyment of them.

THE FREE-RIDER PROBLEM

The citizens of Smalltown, U.S.A., like seeing fireworks on the Fourth of July. Each of the town's 500 residents places a \$10 value on the experience. The cost of putting on a fireworks display is \$1,000. Because the \$5,000 of benefits exceed the \$1,000 of costs, it is efficient for Smalltown residents to see fireworks on the Fourth of July.

Would the private market produce the efficient outcome? Probably not. Imagine that Ellen, a Smalltown entrepreneur, decided to put on a fireworks display. Ellen would surely have trouble selling tickets to the event because her potential customers would quickly figure out that they could see the fireworks even without a ticket. Fireworks are not excludable, so people have an incentive to be free riders. A **free rider** is a person who receives the benefit of a good but avoids paying for it.

free rider

a person who receives the benefit of a good but avoids paying for it

One way to view this market failure is that it arises because of an externality. If Ellen did put on the fireworks display, she would confer an external benefit on those who saw the display without paying for it. When deciding whether to put on the display, Ellen ignores these external benefits. Even though a fireworks display is socially desirable, it is not privately profitable. As a result, Ellen makes the socially inefficient decision not to put on the display.

Although the private market fails to supply the fireworks display demanded by Smalltown residents, the solution to Smalltown's problem is obvious: The local government can sponsor a Fourth of July celebration. The town council can raise everyone's taxes by \$2 and use the revenue to hire Ellen to produce the fireworks. Everyone in Smalltown is better off by \$8—the \$10 in value from the fireworks minus the \$2 tax bill. Ellen can help Smalltown reach the efficient outcome as a public employee even though she could not do so as a private entrepreneur.

The story of Smalltown is simplified, but it is also realistic. In fact, many local governments in the United States do pay for fireworks on the Fourth of July. Moreover, the story shows a general lesson about public goods: Because public goods are not excludable, the free-rider problem prevents the private market from supplying them. The government, however, can potentially remedy the problem. If the government decides that the total benefits exceed the costs, it can provide the public good and pay for it with tax revenue, making everyone better off.

SOME IMPORTANT PUBLIC GOODS

There are many examples of public goods. Here we consider three of the most important.

National Defense The defense of the country from foreign aggressors is a classic example of a public good. It is also one of the most expensive. In 1999 the U.S. federal government spent a total of \$277 billion on national defense, or about \$1,018 per person. People disagree about whether this amount is too small or too large, but almost no one doubts that some government spending for national defense is necessary. Even economists who advocate small government agree that the national defense is a public good the government should provide.

Basic Research The creation of knowledge is a public good. If a mathematician proves a new theorem, the theorem enters the general pool of knowledge



"I like the concept if we can do it with no new taxes."

that anyone can use without charge. Because knowledge is a public good, profit-seeking firms tend to free ride on the knowledge created by others and, as a result, devote too few resources to the creation of knowledge.

In evaluating the appropriate policy toward knowledge creation, it is important to distinguish general knowledge from specific, technological knowledge. Specific, technological knowledge, such as the invention of a better battery, can be patented. The inventor thus obtains much of the benefit of his invention, although certainly not all of it. By contrast, a mathematician cannot patent a theorem; such general knowledge is freely available to everyone. In other words, the patent system makes specific, technological knowledge excludable, whereas general knowledge is not excludable.

The government tries to provide the public good of general knowledge in various ways. Government agencies, such as the National Institutes of Health and the National Science Foundation, subsidize basic research in medicine, mathematics, physics, chemistry, biology, and even economics. Some people justify government funding of the space program on the grounds that it adds to society's pool of knowledge. Certainly, many private goods, including bullet-proof vests and the instant drink Tang, use materials that were first developed by scientists and engineers trying to land a man on the moon. Determining the appropriate level of governmental support for these endeavors is difficult because the benefits are hard to measure. Moreover, the members of Congress who appropriate funds for research usually have little expertise in science and, therefore, are not in the best position to judge what lines of research will produce the largest benefits.

Fighting Poverty Many government programs are aimed at helping the poor. The welfare system (officially called Temporary Assistance for Needy Families) provides a small income for some poor families. Similarly, the Food Stamp program subsidizes the purchase of food for those with low incomes, and various government housing programs make shelter more affordable. These antipoverty programs are financed by taxes on families that are financially more successful.

Economists disagree among themselves about what role the government should play in fighting poverty. Although we will discuss this debate more fully in Chapter 20, here we note one important argument: Advocates of antipoverty programs claim that fighting poverty is a public good.

Suppose that everyone prefers to live in a society without poverty. Even if this preference is strong and widespread, fighting poverty is not a “good” that the private market can provide. No single individual can eliminate poverty because the problem is so large. Moreover, private charity is hard pressed to solve the problem: People who do not donate to charity can free ride on the generosity of others. In this case, taxing the wealthy to raise the living standards of the poor can make everyone better off. The poor are better off because they now enjoy a higher standard of living, and those paying the taxes are better off because they enjoy living in a society with less poverty.

CASE STUDY ARE LIGHTHOUSES PUBLIC GOODS?

Some goods can switch between being public goods and being private goods depending on the circumstances. For example, a fireworks display is a public good if performed in a town with many residents. Yet if performed at a private amusement park, such as Walt Disney World, a fireworks display is more like a private good because visitors to the park pay for admission.

Another example is a lighthouse. Economists have long used lighthouses as examples of a public good. Lighthouses are used to mark specific locations so that passing ships can avoid treacherous waters. The benefit that the lighthouse provides to the ship captain is neither excludable nor rival, so each captain has an incentive to free ride by using the lighthouse to navigate without paying for the service. Because of this free-rider problem, private markets usually fail to provide the lighthouses that ship captains need. As a result, most lighthouses today are operated by the government.



USE OF THE LIGHTHOUSE IS FREE TO THE BOAT OWNER. DOES THIS MAKE THE LIGHTHOUSE A PUBLIC GOOD?

In some cases, however, lighthouses may be closer to private goods. On the coast of England in the nineteenth century, some lighthouses were privately owned and operated. The owner of the local lighthouse did not try to charge ship captains for the service but did charge the owner of the nearby port. If the port owner did not pay, the lighthouse owner turned off the light, and ships avoided that port.

In deciding whether something is a public good, one must determine the number of beneficiaries and whether these beneficiaries can be excluded from enjoying the good. A free-rider problem arises when the number of beneficiaries is large and exclusion of any one of them is impossible. If a lighthouse benefits many ship captains, it is a public good. Yet if it primarily benefits a single port owner, it is more like a private good.

THE DIFFICULT JOB OF COST-BENEFIT ANALYSIS

So far we have seen that the government provides public goods because the private market on its own will not produce an efficient quantity. Yet deciding that the government must play a role is only the first step. The government must then determine what kinds of public goods to provide and in what quantities.

Suppose that the government is considering a public project, such as building a new highway. To judge whether to build the highway, it must compare the total benefits of all those who would use it to the costs of building and maintaining it. To make this decision, the government might hire a team of economists and engineers to conduct a study, called a **cost-benefit analysis**, the goal of which is to estimate the total costs and benefits of the project to society as a whole.

Cost-benefit analysts have a tough job. Because the highway will be available to everyone free of charge, there is no price with which to judge the value of the highway. Simply asking people how much they would value the highway is not reliable. First, quantifying benefits is difficult using the results from a questionnaire. Second, respondents have little incentive to tell the truth. Those who would use the highway have an incentive to exaggerate the benefit they receive to get the highway built. Those who would be harmed by the highway have an incentive to exaggerate the costs to them to prevent the highway from being built.

The efficient provision of public goods is, therefore, intrinsically more difficult than the efficient provision of private goods. Private goods are provided in the market. Buyers of a private good reveal the value they place on it by the prices they are willing to pay. Sellers reveal their costs by the prices they are willing to accept. By contrast, cost-benefit analysts do not observe any price signals when evaluating whether the government should provide a public good. Their findings on the costs and benefits of public projects are, therefore, rough approximations at best.

cost-benefit analysis

a study that compares the costs and benefits to society of providing a public good

CASE STUDY HOW MUCH IS A LIFE WORTH?

Imagine that you have been elected to serve as a member of your local town council. The town engineer comes to you with a proposal: The town can spend \$10,000 to build and operate a traffic light at a town intersection that now has only a stop sign. The benefit of the traffic light is increased safety. The engineer

estimates, based on data from similar intersections, that the traffic light would reduce the risk of a fatal traffic accident over the lifetime of the traffic light from 1.6 to 1.1 percent. Should you spend the money for the new light?

To answer this question, you turn to cost-benefit analysis. But you quickly run into an obstacle: The costs and benefits must be measured in the same units if you are to compare them meaningfully. The cost is measured in dollars, but the benefit—the possibility of saving a person's life—is not directly monetary. To make your decision, you have to put a dollar value on a human life.

At first, you may be tempted to conclude that a human life is priceless. After all, there is probably no amount of money that you could be paid to voluntarily give up your life or that of a loved one. This suggests that a human life has an infinite dollar value.

For the purposes of cost-benefit analysis, however, this answer leads to nonsensical results. If we truly placed an infinite value on human life, we should be placing traffic lights on every street corner. Similarly, we should all be driving large cars with all the latest safety features, instead of smaller ones with fewer safety features. Yet traffic lights are not at every corner, and people sometimes choose to buy small cars without side-impact air bags or antilock brakes. In both our public and private decisions, we are at times willing to risk our lives to save some money.

Once we have accepted the idea that a person's life does have an implicit dollar value, how can we determine what that value is? One approach, sometimes used by courts to award damages in wrongful-death suits, is to look at the total amount of money a person would have earned if he or she had lived. Economists are often critical of this approach. It has the bizarre implication that the life of a retired or disabled person has no value.

A better way to value human life is to look at the risks that people are voluntarily willing to take and how much they must be paid for taking them. Mortality risk varies across jobs, for example. Construction workers in high-rise buildings face greater risk of death on the job than office workers do. By comparing wages in risky and less risky occupations, controlling for education, experience, and other determinants of wages, economists can get some sense about what value people put on their own lives. Studies using this approach conclude that the value of a human life is about \$10 million.

EVERYONE WOULD
LIKE TO AVOID THE
RISK OF THIS, BUT AT
WHAT COST?



IN THE NEWS

Existence Value



COST-BENEFIT ANALYSTS OFTEN RUN INTO hard questions. Here's an example.

They Exist. Therefore They Are. But, Do You Care?

BY SAM HOWE VERHOVEK

It sounds like a philosophical cousin to the age-old question of whether a tree falling in the forest makes a sound if no one is around to hear it. In this case, though, federal officials are seeking to add an economic variable to the puzzle: Just how much is it worth to *you* to know that a once-dammed river is running wild again—even if you never visit it?

In the midst of a major study of whether or not to breach four huge hydroelectric dams on the Snake River in eastern Washington, economists with the Army Corps of Engineers are adding a factor known as “existence value” to their list of costs and benefits of the contentious proposal.

Breaching the dams would restore 140 miles of the lower Snake to its wild, free-flowing condition and would, many biologists argue, stand a good chance of revitalizing endangered salmon runs in the river. Aside from calculating the proposal's effects on jobs, electric bills, and shipping rates, the Government is now hoping to assign a dollar value to Americans' knowledge that a piece of their wilderness might be regained. . . .

“The idea that you'd be willing to pay something for some state of the world to exist, as you would pay for a commodity or a contract for services, is not at all crazy,” said Alan Randall, chairman of the department of agricultural, environmental, and development economics at Ohio State University. “The

controversy, really, is mostly about measurability.”

Proponents of the dam-breaching proposal have pointed to polls suggesting that Seattle-area residents would be willing to pay a few extra dollars a month on their electricity bills in order to save salmon runs. . . . Economists at the Corps of Engineers have calculated that breaching the four Snake River dams and successfully restoring the salmon is an idea for which Americans would be willing to shell out [in total] as much as \$1 billion. . . .

Others question whether such a value can be accurately measured. “The only way to do it is to ask people what they would be willing to pay, and in my view you ask people questions like that and you get very upwardly biased results,” said Jerry Hausman, an economics professor at M.I.T. “When somebody calls you on the phone to ask, it's not real money.”

SOURCE: *The New York Times*, Week in Review, October 17, 1999, p. 5.

We can now return to our original example and respond to the town engineer. The traffic light reduces the risk of fatality by 0.5 percent. Thus, the expected benefit from having the traffic light is $0.005 \times \$10 \text{ million}$, or \$50,000. This estimate of the benefit well exceeds the cost of \$10,000, so you should approve the project.

QUICK QUIZ: What is the *free-rider problem*? ♦ Why does the free-rider problem induce the government to provide public goods? ♦ How should the government decide whether to provide a public good?

COMMON RESOURCES

Common resources, like public goods, are not excludable: They are available free of charge to anyone who wants to use them. Common resources are, however, rival:

Tragedy of the Commons

a parable that illustrates why common resources get used more than is desirable from the standpoint of society as a whole

One person's use of the common resource reduces other people's enjoyment of it. Thus, common resources give rise to a new problem. Once the good is provided, policymakers need to be concerned about how much it is used. This problem is best understood from the classic parable called the **Tragedy of the Commons**.

THE TRAGEDY OF THE COMMONS

Consider life in a small medieval town. Of the many economic activities that take place in the town, one of the most important is raising sheep. Many of the town's families own flocks of sheep and support themselves by selling the sheep's wool, which is used to make clothing.

As our story begins, the sheep spend much of their time grazing on the land surrounding the town, called the Town Common. No family owns the land. Instead, the town residents own the land collectively, and all the residents are allowed to graze their sheep on it. Collective ownership works well because land is plentiful. As long as everyone can get all the good grazing land they want, the Town Common is not a rival good, and allowing residents' sheep to graze for free causes no problems. Everyone in town is happy.

As the years pass, the population of the town grows, and so does the number of sheep grazing on the Town Common. With a growing number of sheep and a fixed amount of land, the land starts to lose its ability to replenish itself. Eventually, the land is grazed so heavily that it becomes barren. With no grass left on the Town Common, raising sheep is impossible, and the town's once prosperous wool industry disappears. Many families lose their source of livelihood.

What causes the tragedy? Why do the shepherds allow the sheep population to grow so large that it destroys the Town Common? The reason is that social and private incentives differ. Avoiding the destruction of the grazing land depends on the collective action of the shepherds. If the shepherds acted together, they could reduce the sheep population to a size that the Town Common can support. Yet no single family has an incentive to reduce the size of its own flock because each flock represents only a small part of the problem.

In essence, the Tragedy of the Commons arises because of an externality. When one family's flock grazes on the common land, it reduces the quality of the land available for other families. Because people neglect this negative externality when deciding how many sheep to own, the result is an excessive number of sheep.

If the tragedy had been foreseen, the town could have solved the problem in various ways. It could have regulated the number of sheep in each family's flock, internalized the externality by taxing sheep, or auctioned off a limited number of sheep-grazing permits. That is, the medieval town could have dealt with the problem of overgrazing in the way that modern society deals with the problem of pollution.

In the case of land, however, there is a simpler solution. The town can divide up the land among town families. Each family can enclose its parcel of land with a fence and then protect it from excessive grazing. In this way, the land becomes a private good rather than a common resource. This outcome in fact occurred during the enclosure movement in England in the seventeenth century.

The Tragedy of the Commons is a story with a general lesson: When one person uses a common resource, he diminishes other people's enjoyment of it. Because of this negative externality, common resources tend to be used excessively.

The government can solve the problem by reducing use of the common resource through regulation or taxes. Alternatively, the government can sometimes turn the common resource into a private good.

This lesson has been known for thousands of years. The ancient Greek philosopher Aristotle pointed out the problem with common resources: “What is common to many is taken least care of, for all men have greater regard for what is their own than for what they possess in common with others.”

SOME IMPORTANT COMMON RESOURCES

There are many examples of common resources. In almost all cases, the same problem arises as in the Tragedy of the Commons: Private decisionmakers use the common resource too much. Governments often regulate behavior or impose fees to mitigate the problem of overuse.

Clean Air and Water As we discussed in Chapter 10, markets do not adequately protect the environment. Pollution is a negative externality that can be remedied with regulations or with Pigovian taxes on polluting activities. One can view this market failure as an example of a common-resource problem. Clean air and clean water are common resources like open grazing land, and excessive pollution is like excessive grazing. Environmental degradation is a modern Tragedy of the Commons.

Oil Pools Consider an underground pool of oil so large that it lies under many properties with different owners. Any of the owners can drill and extract the oil, but when one owner extracts oil, less is available for the others. The oil is a common resource.

Just as the number of sheep grazing on the Town Common was inefficiently large, the number of wells drawing from the oil pool will be inefficiently large. Because each owner who drills a well imposes a negative externality on the other owners, the benefit to society of drilling a well is less than the benefit to the owner who drills it. That is, drilling a well can be privately profitable even when it is socially undesirable. If owners of the properties decide individually how many oil wells to drill, they will drill too many.

To ensure that the oil is extracted at lowest cost, some type of joint action among the owners is necessary to solve the common-resource problem. The Coase theorem, which we discussed in Chapter 10, suggests that a private solution might be possible. The owners could reach an agreement among themselves about how to extract the oil and divide the profits. In essence, the owners would then act as if they were in a single business.

When there are many owners, however, a private solution is more difficult. In this case, government regulation could ensure that the oil is extracted efficiently.

Congested Roads Roads can be either public goods or common resources. If a road is not congested, then one person’s use does not affect anyone else. In this case, use is not rival, and the road is a public good. Yet if a road is congested, then use of that road yields a negative externality. When one person drives on the road, it becomes more crowded, and other people must drive more slowly. In this case, the road is a common resource.

IN THE NEWS*The Singapore Solution*

TOLLS ARE A SIMPLE WAY TO SOLVE THE problem of road congestion and, according to some economists, are not used as much as they should be. In this opinion column, economist Lester Thurow describes Singapore's success in dealing with congestion.

Economics of Road Pricing

BY LESTER C. THUROW

Start with a simple observational truth. No city has ever been able to solve its congestion and pollution problems by building more roads.

Some of the world's cities have built a lot of roads (Los Angeles) and some have very few (Shanghai only recently



HOW CAN WE CLEAR THIS MARKET?

has had a lot of autos) but the degrees of congestion and pollution don't differ very much. More roads simply encourage more people to use their cars, to live farther away from work, and thus use more road space. . . . A recent analysis of congestion problems in London came to the conclusion that London could tear

the entire central city down to make room for roads and would still have something approaching gridlock.

Economists have always had a theoretical answer for auto congestion and pollution problems—road pricing. Charge people for using roads based on what roads they use, what time of day and

One way for the government to address the problem of road congestion is to charge drivers a toll. A toll is, in essence, a Pigovian tax on the externality of congestion. Often, as in the case of local roads, tolls are not a practical solution because the cost of collecting them is too high.

Sometimes congestion is a problem only at certain times of day. If a bridge is heavily traveled only during rush hour, for instance, the congestion externality is larger during this time than during other times of day. The efficient way to deal with these externalities is to charge higher tolls during rush hour. This toll would provide an incentive for drivers to alter their schedules and would reduce traffic when congestion is greatest.

Another policy that responds to the problem of road congestion, discussed in a case study in the previous chapter, is the tax on gasoline. Gasoline is a complementary good to driving: An increase in the price of gasoline tends to reduce the quantity of driving demanded. Therefore, a gasoline tax reduces road congestion.

year they use those roads, and the degree to which pollution problems exist at the time they are using those roads. Set prices at the levels that yield the optimal amounts of usage.

Until Singapore decided to try, no city had ever had the nerve to use road pricing. Many ideas seem good theoretically but have some hidden unexpected flaws. Singapore now has more than a decade of experience. The system works! There are no unexpected flaws. Singapore is the only city on the face of the earth without congestion and auto-induced pollution problems.

In Singapore a series of toll booths surrounds the central core of the city. To drive into the city, each car must pay a toll based on the roads being used, the time of day when the driving will occur, and that day's pollution problem. Prices are raised and lowered to get optimal usage.

In addition, Singapore calculates the maximum number of cars that can be supported without pollution outside of the central city and auctions off the rights to license new cars each month. Different types of plates allow different

degrees of usage. A plate that allows one to use their car at any time is much more expensive than a plate that only allows one to use their car on weekends—a time when congestion problems are much less intense. Prices depend on supply and demand.

With this system Singapore ends up not wasting resources on infrastructure projects that won't cure congestion and pollution problems. The revenue collected from the system is used to lower other taxes.

If that is so, why then did London reject road pricing in its recent report on its auto congestion and pollution problems? They feared that such a system would be seen as too much interference from the heavy hand of government and that the public would not put up with a system that allows the rich to drive more than the poor.

Both arguments ignore the fact that we already have toll roads, but new technologies now also make it possible to avoid both problems.

Using bar codes and debit cards, a city can install bar code readers at different points around the city. As any car

goes by each point a certain amount is deducted from the driver's debit card account depending upon weather, time of day, and location.

Inside the car, the driver has a meter that tells him how much he has been charged and how much remains in his debit card account. . . .

If one is an egalitarian and thinks that driving privileges should be distributed equally (i.e., not based upon income) then each auto can be given a specified debit card balance every year and those who are willing to drive less can sell their unused balances to those that want to drive more.

Instead of giving the city extra tax revenue, this system gives those who are willing to live near work or to use public transit an income supplement. Since poor people drive less than rich people, the system ends up being an egalitarian redistribution of income from the rich to the poor.

SOURCE: *The Boston Globe*, February 28, 1995, p. 40.

A gasoline tax, however, is an imperfect solution to road congestion. The problem is that the gasoline tax affects other decisions besides the amount of driving on congested roads. For example, the gasoline tax discourages driving on noncongested roads, even though there is no congestion externality for these roads.

Fish, Whales, and Other Wildlife Many species of animals are common resources. Fish and whales, for instance, have commercial value, and anyone can go to the ocean and catch whatever is available. Each person has little incentive to maintain the species for the next year. Just as excessive grazing can destroy the Town Common, excessive fishing and whaling can destroy commercially valuable marine populations.

The ocean remains one of the least regulated common resources. Two problems prevent an easy solution. First, many countries have access to the oceans, so any solution would require international cooperation among countries that hold

IN THE NEWS*Should Yellowstone Charge as Much as Disney World?*

NATIONAL PARKS, LIKE ROADS, CAN BE either public goods or common resources. If congestion is not a problem, a visit to a park is not rival. Yet once a park becomes popular, it suffers from the same problem as the Town Common. In this opinion column, an economist argues for the use of higher entrance fees to solve the problem.

Save the Parks, and Make a Profit

BY ALLEN R. SANDERSON

It is common knowledge that our national parks are overcrowded, deteriorating, and broke. Some suggest that we address these problems by requiring reservations, closing some areas, or asking

Congress to increase financing to the National Park Service. But to an economist, there is a more obvious solution: Raise the entrance fees.

When the National Park Service was established in 1916, the admission price to Yellowstone for a family of five arriving by car was \$7.50; today, the price is only \$10. Had the 1916 price been adjusted for inflation, the comparable 1995 fee would be \$120 a day—about what that family would pay for a day of rides at Disney World, . . . or to see a professional football game.

No wonder our national parks are overrun and overtrampled. We are treating our natural and historical treasures as free goods when they are not. We are ignoring the costs of maintaining these places and rationing by congestion—when it gets too crowded, no more visitors are allowed—perhaps the most inefficient way to allocate scarce resources. The price of a family's day in a national park has not kept pace with most other forms of recreation. Systemwide, it barely averages a dollar a person. . . .

An increase in daily user fees to, say, \$20 per person would either reduce

the overcrowding and deterioration in our parks by cutting down on the number of visitors or it would substantially raise fee revenues for the Park Service (assuming that legislation was passed that would let the park system keep this money). Greater revenue is the more likely outcome. After spending several hundred dollars to reach Yellowstone Park, few people would be deterred by another \$20.

The added revenues would bring more possibilities for outdoor recreation, both through expansion of the National Park Service and by encouraging private entrepreneurs to carve out and operate their own parks, something they cannot do alongside a public competitor giving away his product well below cost.

It is time to put our money where our Patagonia outfits are: Either we value the Grand Canyon and Yosemite and won't complain about paying a realistic entrance fee, or we don't really value them and shouldn't wring our hands over their present sorry state and likely sorrier fate.

SOURCE: *The New York Times*, September 30, 1995, p. 19.

different values. Second, because the oceans are so vast, enforcing any agreement is difficult. As a result, fishing rights have been a frequent source of international tension among normally friendly countries.

Within the United States, various laws aim to protect fish and other wildlife. For example, the government charges for fishing and hunting licenses, and it restricts the lengths of the fishing and hunting seasons. Fishermen are often required to throw back small fish, and hunters can kill only a limited number of animals. All these laws reduce the use of a common resource and help maintain animal populations.

CASE STUDY WHY THE COW IS NOT EXTINCT

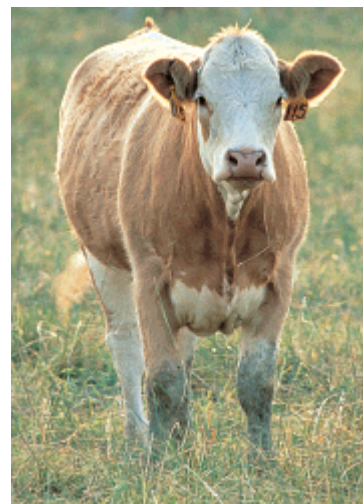
Throughout history, many species of animals have been threatened with extinction. When Europeans first arrived in North America, more than 60 million

buffalo roamed the continent. Yet hunting the buffalo was so popular during the nineteenth century that by 1900 the animal's population fell to about 400 before the government stepped in to protect the species. In some African countries today, the elephant faces a similar challenge, as poachers kill the animals for the ivory in their tusks.

Yet not all animals with commercial value face this threat. The cow, for example, is a valuable source of food, but no one worries that the cow will soon be extinct. Indeed, the great demand for beef seems to ensure that the species will continue to thrive.

Why is the commercial value of ivory a threat to the elephant, while the commercial value of beef is a guardian of the cow? The reason is that elephants are a common resource, whereas cows are a private good. Elephants roam freely without any owners. Each poacher has a strong incentive to kill as many elephants as he can find. Because poachers are numerous, each poacher has only a slight incentive to preserve the elephant population. By contrast, cows live on ranches that are privately owned. Each rancher takes great effort to maintain the cow population on his ranch because he reaps the benefit of these efforts.

Governments have tried to solve the elephant's problem in two ways. Some countries, such as Kenya, Tanzania, and Uganda, have made it illegal to kill elephants and sell their ivory. Yet these laws have been hard to enforce, and elephant populations have continued to dwindle. By contrast, other countries, such as Botswana, Malawi, Namibia, and Zimbabwe, have made elephants a private good by allowing people to kill elephants, but only those on their own property. Landowners now have an incentive to preserve the species on their own land, and as a result, elephant populations have started to rise. With private ownership and the profit motive now on its side, the African elephant might someday be as safe from extinction as the cow.



"WILL THE MARKET PROTECT ME?"

QUICK QUIZ: Why do governments try to limit the use of common resources?

CONCLUSION: THE IMPORTANCE OF PROPERTY RIGHTS

In this chapter and the previous one, we have seen there are some "goods" that the market does not provide adequately. Markets do not ensure that the air we breathe is clean or that our country is defended from foreign aggressors. Instead, societies rely on the government to protect the environment and to provide for the national defense.

Although the problems we considered in these chapters arise in many different markets, they share a common theme. In all cases, the market fails to allocate resources efficiently because *property rights* are not well established. That is, some item of value does not have an owner with the legal authority to control it. For example, although no one doubts that the "good" of clean air or national defense is valuable, no one has the right to attach a price to it and profit from its use. A factory

pollutes too much because no one charges the factory for the pollution it emits. The market does not provide for national defense because no one can charge those who are defended for the benefit they receive.

When the absence of property rights causes a market failure, the government can potentially solve the problem. Sometimes, as in the sale of pollution permits, the solution is for the government to help define property rights and thereby unleash market forces. Other times, as in the restriction on hunting seasons, the solution is for the government to regulate private behavior. Still other times, as in the provision of national defense, the solution is for the government to supply a good that the market fails to supply. In all cases, if the policy is well planned and well run, it can make the allocation of resources more efficient and thus raise economic well-being.

Summary

- ◆ Goods differ in whether they are excludable and whether they are rival. A good is excludable if it is possible to prevent someone from using it. A good is rival if one person's enjoyment of the good prevents other people from enjoying the same unit of the good. Markets work best for private goods, which are both excludable and rival. Markets do not work as well for other types of goods.
- ◆ Public goods are neither rival nor excludable. Examples of public goods include fireworks displays, national defense, and the creation of fundamental knowledge. Because people are not charged for their use of the public good, they have an incentive to free ride when the good is provided privately. Therefore, governments provide public goods, making their decision about the quantity based on cost-benefit analysis.
- ◆ Common resources are rival but not excludable. Examples include common grazing land, clean air, and congested roads. Because people are not charged for their use of common resources, they tend to use them excessively. Therefore, governments try to limit the use of common resources.

Key Concepts

excludability, p. 226
rivalry, p. 226
private goods, p. 226

public goods, p. 226
common resources, p. 226
free rider, p. 228

cost-benefit analysis, p. 231
Tragedy of the Commons, p. 234

Questions for Review

1. Explain what is meant by a good being "excludable." Explain what is meant by a good being "rival." Is a pizza excludable? Is it rival?
2. Define and give an example of a public good. Can the private market provide this good on its own? Explain.
3. What is cost-benefit analysis of public goods? Why is it important? Why is it hard?
4. Define and give an example of a common resource. Without government intervention, will people use this good too much or too little? Why?

Problems and Applications

1. The text says that both public goods and common resources involve externalities.
 - a. Are the externalities associated with public goods generally positive or negative? Use examples in your answer. Is the free-market quantity of public goods generally greater or less than the efficient quantity?
 - b. Are the externalities associated with common resources generally positive or negative? Use examples in your answer. Is the free-market use of common resources generally greater or less than the efficient use?
2. Think about the goods and services provided by your local government.
 - a. Using the classification in Figure 11-1, explain what category each of the following goods falls into:
 - ◆ police protection
 - ◆ snow plowing
 - ◆ education
 - ◆ rural roads
 - ◆ city streets
 - b. Why do you think the government provides items that are not public goods?
3. Charlie loves watching *Teletubbies* on his local public TV station, but he never sends any money to support the station during their fund-raising drives.
 - a. What name do economists have for Charlie?
 - b. How can the government solve the problem caused by people like Charlie?
 - c. Can you think of ways the private market can solve this problem? How does the existence of cable TV alter the situation?
4. The text states that private firms will not undertake the efficient amount of basic scientific research.
 - a. Explain why this is so. In your answer, classify basic research in one of the categories shown in Figure 11-1.
 - b. What sort of policy has the United States adopted in response to this problem?
 - c. It is often argued that this policy increases the technological capability of American producers relative to that of foreign firms. Is this argument consistent with your classification of basic research in part (a)? (Hint: Can excludability apply to some potential beneficiaries of a public good and not others?)
5. Why is there litter along most highways but rarely in people's yards?
6. The Washington, D.C., metro (subway) system charges higher fares during rush hours than during the rest of the day. Why might it do this?
7. Timber companies in the United States cut down many trees on publicly owned land and many trees on privately owned land. Discuss the likely efficiency of logging on each type of land in the absence of government regulation. How do you think the government should regulate logging on publicly owned lands? Should similar regulations apply to privately owned land?
8. An *Economist* article (March 19, 1994) states: "In the past decade, most of the rich world's fisheries have been exploited to the point of near-exhaustion." The article continues with an analysis of the problem and a discussion of possible private and government solutions:
 - a. "Do not blame fishermen for overfishing. They are behaving rationally, as they have always done." In what sense is "overfishing" rational for fishermen?
 - b. "A community, held together by ties of obligation and mutual self-interest, can manage a common resource on its own." Explain how such management can work in principle, and what obstacles it faces in the real world.
 - c. "Until 1976 most world fish stocks were open to all comers, making conservation almost impossible. Then an international agreement extended some aspects of [national] jurisdiction from 12 to 200 miles offshore." Using the concept of property rights, discuss how this agreement reduces the scope of the problem.
 - d. The article notes that many governments come to the aid of suffering fishermen in ways that encourage increased fishing. How do such policies encourage a vicious cycle of overfishing?
 - e. "Only when fishermen believe they are assured a long-term and exclusive right to a fishery are they likely to manage it in the same far-sighted way as good farmers manage their land." Defend this statement.
 - f. What other policies to reduce overfishing might be considered?
9. In a market economy, information about the quality or function of goods and services is a valuable good in its

own right. How does the private market provide this information? Can you think of any way in which the government plays a role in providing this information?

10. Do you think the Internet is a public good? Why or why not?
11. High-income people are willing to pay more than lower-income people to avoid the risk of death. For example,

they are more likely to pay for safety features on cars. Do you think cost-benefit analysts should take this fact into account when evaluating public projects? Consider, for instance, a rich town and a poor town, both of which are considering the installation of a traffic light. Should the rich town use a higher dollar value for a human life in making this decision? Why or why not?

12



THE DESIGN OF THE TAX SYSTEM

Al “Scarface” Capone, the notorious 1920s gangster and crime boss, was never convicted for his many violent crimes. Yet eventually he did go to jail—for tax evasion. He had neglected to heed Ben Franklin’s observation that “in this world nothing is certain, but death and taxes.”

When Franklin made this claim in 1789, the average American paid less than 5 percent of his income in taxes, and that remained true for the next hundred years. Over the course of the twentieth century, however, taxes have become ever more important in the life of the typical person. Today, all taxes taken together—including personal income taxes, corporate income taxes, payroll taxes, sales taxes, and property taxes—use up about a third of the average American’s income. In many European countries, the tax bite is even larger.

Taxes are inevitable because we as citizens expect the government to provide us with various goods and services. The previous two chapters have started to

IN THIS CHAPTER YOU WILL . . .

Get an overview of how the U.S. government raises and spends money

Examine the efficiency costs of taxes

Learn alternative ways to judge the equity of a tax system

See why studying tax incidence is crucial for evaluating tax equity

Consider the tradeoff between efficiency and equity in the design of a tax system



shed light on one of the *Ten Principles of Economics* from Chapter 1: The government can sometimes improve market outcomes. When the government remedies an externality (such as air pollution), provides a public good (such as national defense), or regulates the use of a common resource (such as fish in a public lake), it can raise economic well-being. Yet the benefits of government come with costs. For the government to perform these and its many other functions, it needs to raise revenue through taxation.

We began our study of taxation in earlier chapters, where we saw how a tax on a good affects supply and demand for that good. In Chapter 6 we saw that a tax reduces the quantity sold in a market, and we examined how the burden of a tax is shared by buyers and sellers, depending on the elasticities of supply and demand. In Chapter 8 we examined how taxes affect economic well-being. We learned that taxes cause *deadweight losses*: The reduction in consumer and producer surplus resulting from a tax exceeds the revenue raised by the government.

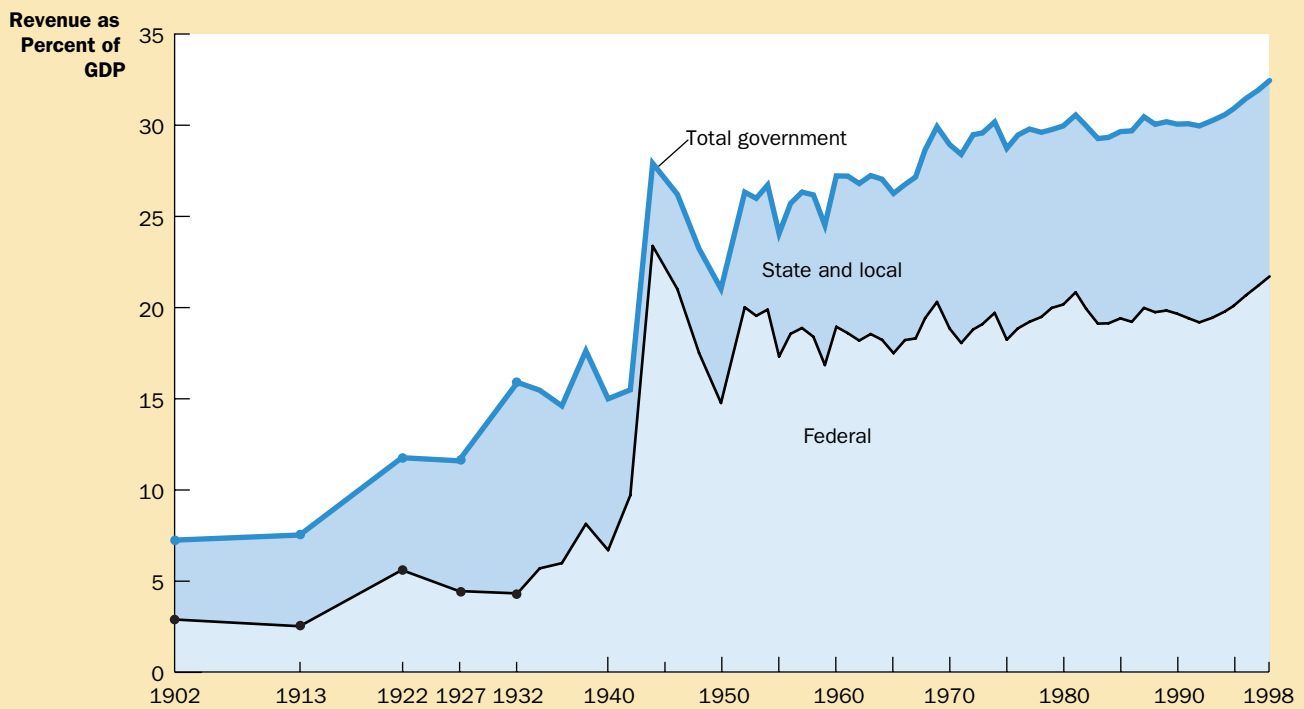
In this chapter we build on these lessons to discuss the design of a tax system. We begin with a financial overview of the U.S. government. When thinking about the tax system, it is useful to know some basic facts about how the U.S. government raises and spends money. We then consider the fundamental principles of taxation. Most people agree that taxes should impose as small a cost on society as possible and that the burden of taxes should be distributed fairly. That is, the tax system should be both *efficient* and *equitable*. As we will see, however, stating these goals is easier than achieving them.

A FINANCIAL OVERVIEW OF THE U.S. GOVERNMENT

How much of the nation's income does the government take as taxes? Figure 12-1 shows government revenue, including federal, state, and local governments, as a percentage of total income for the U.S. economy. It shows that, over time, the government has taken a larger and larger share of total income. In 1902, the government collected 7 percent of total income; in 1998, it collected 32 percent. In other words, as the economy's income has grown, the government has grown even more.

Table 12-1 compares the tax burden for several major countries, as measured by the central government's tax revenue as a percentage of the nation's total income. The United States is in the middle of the pack. The U.S. tax burden is low compared to many European countries, but it is high compared to many other nations around the world. Poor countries, such as India and Pakistan, usually have relatively low tax burdens. This fact is consistent with the evidence in Figure 12-1 of a growing tax burden over time: As a nation gets richer, the government typically takes a larger share of income in taxes.

The overall size of government tells only part of the story. Behind the total dollar figures lie thousands of individual decisions about taxes and spending. To understand the government's finances more fully, let's look at how the total breaks down into some broad categories.



GOVERNMENT REVENUE AS A PERCENTAGE OF GDP. This figure shows revenue of the federal government and of state and local governments as a percentage of gross domestic product (GDP), which measures total income in the economy. It shows that the government plays a large role in the U.S. economy and that its role has grown over time.

SOURCE: *Historical Statistics of the United States*; *Economic Report of the President 1999*; and author's calculations.

Figure 12-1

France	38.8%
United Kingdom	33.7
Germany	29.4
Brazil	19.7
United States	19.3
Canada	18.5
Russia	17.4
Pakistan	15.3
Indonesia	14.7
Mexico	12.8
India	10.3

SOURCE: *World Development Report 1998/99*.

Table 12-1

CENTRAL GOVERNMENT TAX
REVENUE AS A PERCENT OF GDP

THE FEDERAL GOVERNMENT

The U.S. federal government collects about two-thirds of the taxes in our economy. It raises this money in a number of ways, and it finds even more ways to spend it.

Receipts Table 12-2 shows the receipts of the federal government in 1999. Total receipts in this year were \$1,806 billion, a number so large that it is hard to comprehend. To bring this astronomical number down to earth, we can divide it by the size of the U.S. population, which was about 272 million in 1999. We then find that the average American paid \$6,639 to the federal government. A typical family of four paid \$26,556.

The largest source of revenue for the federal government is the individual income tax. As April 15 approaches, almost every American family fills out a tax form to determine how much income tax it owes the government. Each family is required to report its income from all sources: wages from working, interest on savings, dividends from corporations in which it owns shares, profits from any small businesses it operates, and so on. The family's *tax liability* (how much it owes) is then based on its total income.

A family's income tax liability is not simply proportional to its income. Instead, the law requires a more complicated calculation. Taxable income is computed as total income minus an amount based on the number of dependents (primarily children) and minus certain expenses that policymakers have deemed "deductible" (such as mortgage interest payments and charitable giving). Then the tax liability is calculated from taxable income using the schedule shown in Table 12-3. This table presents the *marginal tax rate*—the tax rate applied to each additional dollar of income. Because the marginal tax rate rises as income rises, higher-income families pay a larger percentage of their income in taxes. (We discuss the concept of marginal tax rate more fully later in this chapter.)

Almost as important to the federal government as the individual income tax are payroll taxes. A *payroll tax* is a tax on the wages that a firm pays its workers. Table 12-2 calls this revenue *social insurance taxes* because the revenue from these taxes is earmarked to pay for Social Security and Medicare. Social Security is an income support program, designed primarily to maintain the living standards of the elderly. Medicare is the government health program for the elderly. Table 12-2 shows that the average American paid \$2,239 in social insurance taxes in 1999.

Table 12-2

RECEIPTS OF THE FEDERAL
GOVERNMENT: 1999

TAX	AMOUNT (IN BILLIONS)	AMOUNT PER PERSON	PERCENT OF RECEIPTS
Individual income taxes	\$ 869	\$3,194	48%
Social insurance taxes	609	2,239	34
Corporate income taxes	182	669	10
Other	146	537	8
Total	\$1,806	\$6,639	100%

SOURCE: *Economic Report of the President*, 1999, table B-80.

ON TAXABLE INCOME . . .

THE TAX RATE IS . . .

Up to \$25,750	15.0%
From \$25,750 to \$62,450	28.0
From \$62,450 to \$130,250	31.0
From \$130,250 to \$283,150	36.0
Over \$283,150	39.6

Table 12-3

THE FEDERAL INCOME TAX RATES: 1999. This table shows the marginal tax rates for an unmarried taxpayer. The taxes owed by a taxpayer depend on all the marginal tax rates up to his or her income level. For example, a taxpayer with income of \$50,000 pays 15 percent of the first \$25,750 of income, and then 28 percent of the rest.

Next in magnitude, but much smaller than either individual income taxes or social insurance taxes, is the corporate income tax. A *corporation* is a business that is set up as a separate legal entity. The government taxes each corporation based on its profit—the amount the corporation receives for the goods or services it sells minus the costs of producing those goods or services. Notice that corporate profits are, in essence, taxed twice. They are taxed once by the corporate income tax when the corporation earns the profits; they are taxed a second time by the individual income tax when the corporation uses its profits to pay dividends to its shareholders.

The last category, labeled “other” in Table 12-2, makes up 8 percent of receipts. This category includes *excise taxes*, which are taxes on specific goods like gasoline, cigarettes, and alcoholic beverages. It also includes various small items, such as estate taxes and customs duties.

Spending Table 12-4 shows the spending of the federal government in 1999. Total spending was \$1,727 billion, or \$6,350 per person. This table also shows how the federal government’s spending was divided among major categories.

The largest category in Table 12-4 is Social Security, which represents mostly transfer payments to the elderly. (A *transfer payment* is a government payment not made in exchange for a good or service.) This category made up 23 percent of spending by the federal government in 1999 and is growing in importance. The reason for its growth is that increases in life expectancy and decreases in birthrates have caused the elderly population to grow more rapidly than the total population. Most analysts expect this trend to continue for many years into the future.

The second largest category of spending is national defense. This includes both the salaries of military personnel and the purchases of military equipment such as guns, fighter jets, and warships. Spending on national defense fluctuates over time as international tensions and the political climate change. Not surprisingly, spending on national defense rises substantially during wars.

The third category is spending on income security, which includes transfer payments to poor families. One program is Temporary Assistance for Needy Families (TANF), often simply called “welfare.” Another is the Food Stamp program, which gives poor families vouchers that they can use to buy food. The federal government pays some of this money to state and local governments, which administer the programs under federal guidelines.

A bit smaller than spending on income security is net interest. When a person borrows from a bank, the bank requires the borrower to pay interest for the loan. The same is true when the government borrows from the public. The more indebted the government, the larger the amount it must spend in interest payments.

Table 12-4SPENDING OF THE FEDERAL
GOVERNMENT: 1999

CATEGORY	AMOUNT (IN BILLIONS)	AMOUNT PER PERSON	PERCENT OF SPENDING
Social Security	\$ 393	\$1,445	23%
National defense	277	1,018	16
Income security	243	893	14
Net interest	227	835	13
Medicare	205	754	12
Health	143	526	8
Other	239	879	14
Total	\$1,727	\$6,350	100%

SOURCE: *Economic Report of the President*, 1999, table B-80.

Medicare, the next category in Table 12-4, is the government's health plan for the elderly. Spending in this category has risen substantially over time for two reasons. First, the elderly population has grown more quickly than the overall population. Second, the cost of health care has risen more rapidly than the cost of other goods and services. The rapid growth of this budget item is one reason that President Clinton and others have proposed reforms of the health care system.

The next category is health spending other than Medicare. This includes Medicaid, the federal health program for the poor. It also includes spending on medical research, such as through the National Institutes of Health.

The "other" category in Table 12-4 consists of many less expensive functions of government. It includes, for example, the federal court system, the space program, and farm-support programs, as well as the salaries of Congress and the president.

You might have noticed that total receipts of the federal government shown in Table 12-2 exceed total spending shown in Table 12-4 by \$79 billion. Such an excess of receipts over spending is called a **budget surplus**. When receipts fall short of spending, the government is said to run a **budget deficit**. The government finances the budget deficit by borrowing from the public. When the government runs a budget surplus, it uses the excess receipts to reduce its outstanding debts.

budget surplus

*an excess of government receipts
over government spending*

budget deficit

*an excess of government spending
over government receipts*

STATE AND LOCAL GOVERNMENT

State and local governments collect about 40 percent of all taxes paid. Let's look at how they obtain tax revenue and how they spend it.

Receipts Table 12-5 shows the receipts of U.S. state and local governments. Total receipts for 1996 were \$1,223 billion. Based on the 1996 population of about 265 million, this equals \$4,615 per person. The table also shows how this total is broken down into different kinds of taxes.

The two most important taxes for state and local governments are sales taxes and property taxes. Sales taxes are levied as a percentage of the total amount spent at retail stores. Every time a customer buys something, he or she pays the

Table 12-5

RECEIPTS OF STATE AND LOCAL GOVERNMENTS: 1996

TAX	AMOUNT (IN BILLIONS)	AMOUNT PER PERSON	PERCENT OF OF RECEIPTS
Sales taxes	\$ 249	\$ 940	20%
Property taxes	209	789	17
Individual income taxes	147	554	12
Corporate income taxes	32	121	3
From federal government	235	887	19
Other	351	1,324	29
Total	\$1,223	\$4,615	100%

SOURCE: *Economic Report of the President*, 1999, table B-86.

storekeeper an extra amount that the storekeeper remits to the government. (Some states exclude certain items that are considered necessities, such as food and clothing.) Property taxes are levied as a percentage of the estimated value of land and structures, and are paid by property owners. Together these two taxes make up more than a third of all receipts of state and local governments.

State and local governments also levy individual and corporate income taxes. In many cases, state and local income taxes are similar to federal income taxes. In other cases, they are quite different. For example, some states tax income from wages less heavily than income earned in the form of interest and dividends. Some states do not tax income at all.

State and local governments also receive substantial funds from the federal government. To some extent, the federal government's policy of sharing its revenue with state governments redistributes funds from high-income states (who pay more taxes) to low-income states (who receive more benefits). Often these funds are tied to specific programs that the federal government wants to subsidize.

Finally, state and local governments receive much of their receipts from various sources included in the "other" category in Table 12-5. These include fees for fishing and hunting licenses, tolls from roads and bridges, and fares for public buses and subways.

Spending Table 12-6 shows the total spending of state and local governments in 1996 and its breakdown among the major categories.

By far the biggest single expenditure for state and local governments is education. Local governments pay for the public schools, which educate most students from kindergarten to high school. State governments contribute to the support of public universities. In 1996, education accounted for a third of the spending of state and local governments.

The second largest category of spending is for public welfare, which includes transfer payments to the poor. This category includes some federal programs that are administered by state and local governments. The next category is highways, which includes the building of new roads and the maintenance of existing ones. The "other" category in Table 12-6 includes the many additional services provided by state and local governments, such as libraries, police, garbage removal, fire protection, park maintenance, and snow removal.

Table 12-6

SPENDING OF STATE AND LOCAL GOVERNMENTS: 1996

CATEGORY	AMOUNT (IN BILLIONS)	AMOUNT PER PERSON	PERCENT OF SPENDING
Education	\$ 399	\$1,506	33%
Public welfare	197	743	17
Highways	79	298	7
Other	518	1,955	43
Total	\$1,193	\$4,502	100%

SOURCE: *Economic Report of the President*, 1999, table B-86.

QUICK QUIZ: What are the two most important sources of tax revenue for the federal government? ♦ What are the two most important sources of tax revenue for state and local governments?

TAXES AND EFFICIENCY

Now that we have seen how the U.S. government at various levels raises and spends money, let's consider how one might evaluate its tax policy. Obviously, the aim of a tax system is to raise revenue for the government. But there are many ways to raise any given amount of money. In designing a tax system, policymakers have two objectives: efficiency and equity.

One tax system is more efficient than another if it raises the same amount of revenue at a smaller cost to taxpayers. What are the costs of taxes to taxpayers? The most obvious cost is the tax payment itself. This transfer of money from the taxpayer to the government is an inevitable feature of any tax system. Yet taxes also impose two other costs, which well-designed tax policy tries to avoid or, at least, minimize:

- ♦ The deadweight losses that result when taxes distort the decisions that people make
- ♦ The administrative burdens that taxpayers bear as they comply with the tax laws

An efficient tax system is one that imposes small deadweight losses and small administrative burdens.

DEADWEIGHT LOSSES

Taxes affect the decisions that people make. If the government taxes ice cream, people eat less ice cream and more frozen yogurt. If the government taxes housing, people live in smaller houses and spend more of their income on other things. If the government taxes labor earnings, people work less and enjoy more leisure.

Because taxes distort incentives, they entail deadweight losses. As we first discussed in Chapter 8, the deadweight loss of a tax is the reduction in economic well-being of taxpayers in excess of the amount of revenue raised by the government. The deadweight loss is the inefficiency that a tax creates as people allocate resources according to the tax incentive rather than the true costs and benefits of the goods and services that they buy and sell.

To recall how taxes cause deadweight losses, consider an example. Suppose that Joe places an \$8 value on a pizza, and Jane places a \$6 value on it. If there is no tax on pizza, the price of pizza will reflect the cost of making it. Let's suppose that the price of pizza is \$5, so both Joe and Jane choose to buy one. Both consumers get some surplus of value over the amount paid. Joe gets consumer surplus of \$3, and Jane gets consumer surplus of \$1. Total surplus is \$4.

Now suppose that the government levies a \$2 tax on pizza and the price of pizza rises to \$7. Joe still buys a pizza, but now he has consumer surplus of only \$1. Jane now decides not to buy a pizza because its price is higher than its value to her. The government collects tax revenue of \$2 on Joe's pizza. Total consumer surplus has fallen by \$3 (from \$4 to \$1). Because total surplus has fallen by more than the tax revenue, the tax has a deadweight loss. In this case, the deadweight loss is \$1.

Notice that the deadweight loss comes not from Joe, the person who pays the tax, but from Jane, the person who doesn't. The reduction of \$2 in Joe's surplus exactly offsets the amount of revenue the government collects. The deadweight loss arises because the tax causes Jane to alter her behavior. When the tax raises the price of pizza, Jane is worse off, and yet there is no offsetting revenue to the government. This reduction in Jane's welfare is the deadweight loss of the tax.



"I was gonna fix the place up, but if I did the city would just raise my taxes!"

CASE STUDY SHOULD INCOME OR CONSUMPTION BE TAXED?

When taxes induce people to change their behavior—such as inducing Jane to buy less pizza—the taxes cause deadweight losses and make the allocation of resources less efficient. As we have already seen, much government revenue comes from the individual income tax. In a case study in Chapter 8, we discussed how this tax discourages people from working as hard as they otherwise might. Another inefficiency caused by this tax is that it discourages people from saving.

Consider a person 25 years old who is considering saving \$100. If he puts this money in a savings account that earns 8 percent and leaves it there, he would have \$2,172 when he retires at age 65. Yet if the government taxes one-fourth of his interest income each year, the effective interest rate is only 6 percent. After 40 years of earning 6 percent, the \$100 grows to only \$1,029, less than half of what it would have been without taxation. Thus, because interest income is taxed, saving is much less attractive.

Some economists advocate eliminating the current tax system's disincentive toward saving by changing the basis of taxation. Rather than taxing the amount of income that people *earn*, the government could tax the amount that people *spend*. Under this proposal, all income that is saved would not be taxed until the saving is later spent. This alternative system, called a *consumption tax*, would not distort people's saving decisions.

This idea has some support from policymakers. Representative Bill Archer, who in 1995 became chairman of the powerful House Ways and Means

Committee, has advocated replacing the current income tax system with a consumption tax. Moreover, various provisions of the current tax code already make the tax system a bit like a consumption tax. Taxpayers can put a limited amount of their saving into special accounts—such as Individual Retirement Accounts, Keogh plans, and 401(k) plans—that escape taxation until the money is withdrawn at retirement. For people who do most of their saving through these retirement accounts, their tax bill is, in effect, based on their consumption rather than their income.

ADMINISTRATIVE BURDEN

If you ask the typical person on April 15 for an opinion about the tax system, you might hear about the headache of filling out tax forms. The administrative burden of any tax system is part of the inefficiency it creates. This burden includes not only the time spent in early April filling out forms but also the time spent throughout the year keeping records for tax purposes and the resources the government has to use to enforce the tax laws.

Many taxpayers—especially those in higher tax brackets—hire tax lawyers and accountants to help them with their taxes. These experts in the complex tax laws fill out the tax forms for their clients and help clients arrange their affairs in a way that reduces the amount of taxes owed. This behavior is legal tax avoidance, which is different from illegal tax evasion.

Critics of our tax system say that these advisers help their clients avoid taxes by abusing some of the detailed provisions of the tax code, often dubbed “loopholes.” In some cases, loopholes are congressional mistakes: They arise from ambiguities or omissions in the tax laws. More often, they arise because Congress has chosen to give special treatment to specific types of behavior. For example, the U.S. federal tax code gives preferential treatment to investors in municipal bonds because Congress wanted to make it easier for state and local governments to borrow money. To some extent, this provision benefits states and localities; to some extent, it benefits high-income taxpayers. Most loopholes are well known by those in Congress who make tax policy, but what looks like a loophole to one taxpayer may look like a justifiable tax deduction to another.

The resources devoted to complying with the tax laws are a type of dead-weight loss. The government gets only the amount of taxes paid. By contrast, the taxpayer loses not only this amount but also the time and money spent documenting, computing, and avoiding taxes.

The administrative burden of the tax system could be reduced by simplifying the tax laws. Yet simplification is often politically difficult. Most people are ready to simplify the tax code by eliminating the loopholes that benefit others, yet few are eager to give up the loopholes that they use. In the end, the complexity of the tax law results from the political process as various taxpayers with their own special interests lobby for their causes.

average tax rate

total taxes paid divided by total income

marginal tax rate

the extra taxes paid on an additional dollar of income

MARGINAL TAX RATES VERSUS AVERAGE TAX RATES

When discussing the efficiency and equity of income taxes, economists distinguish between two notions of the tax rate: the average and the marginal. The **average tax rate** is total taxes paid divided by total income. The **marginal tax rate** is the extra taxes paid on an additional dollar of income.

IN THE NEWS*Small Business and
the Tax Laws*

PEOPLE RUNNING SMALL BUSINESSES ARE most aware of the administrative burden of the tax system. Small firms must comply with many of the same laws as large ones. Yet, because of their size, compliance can take a much larger fraction of their revenue. According to one study, the administrative burden is ten times larger for small firms than for large firms. The following article describes some of these costs.

**Obeying the Tax Laws:
Small Business's Burden**

BY ROBERT D. HERSHEY, JR.

In the grand scheme of a federal system that collects more than \$1 trillion a year, Dante's Restaurant, Inc., a modest three-city chain in Pennsylvania, counts for little.

But to people like Lewis Kamin, Dante's controller, the Internal Revenue Code is a year-round headache. There are the biweekly remittances of Social Security and withheld income tax, quarterly reports of payroll and unemployment taxes, quarterly estimated corporate income taxes, and, of course, the maintaining of various records,

including tips, W-4s for withholding, and I-9 citizenship forms.

All this doesn't count the ubiquitous state and local levies that in Dante's case are complicated because liquor-license considerations mean each of its ten restaurants must be separately incorporated.

"There is a lot to watch, a lot to worry about," Mr. Kamin grumbled. . . .

This is the real-world side of American taxation, the federal chunk of which is a system based on a monumentally complex set of laws and regulations that was just one-third its current size when Jimmy Carter called it "a disgrace to the human race."

The code is administered by a 115,000-member Internal Revenue Service army with a \$7 billion budget. But that amount is dwarfed by what taxpayers themselves spend on meeting their obligations.

Estimates of what it costs American businesses to comply with federal tax law reach into the hundreds of billions of dollars a year. . . . Big companies are under almost continuous audit. The 1992 return for one giant company ran to 21,000 pages and 30 volumes. But the heaviest burden by far falls on small business.

In fact, according to Arthur P. Hall, a senior fellow at The Tax Foundation, the local hardware store, delicatessen, or gas station with assets of less than \$1 million—a category embracing 90 percent of the nation's corporations—spends \$390 for each \$100 it sends to Washington. Put another way, the government got just \$4.1 billion from these

businesses in 1990, compared with the \$15.9 billion they spent producing the basic corporate forms, the 1120 and 1120S.

"What this means is that the corporate income tax is a very inefficient revenue source for the federal government," Mr. Hall said. . . .

Although complaints about the tax system are often aimed at the I.R.S., businesspeople and policymakers generally contend that the real fault lies with Congress and its frequent, often well-intentioned tinkering with the law. The resulting complexity is taking an ever-mounting toll on respect for the system, and thereby undermines the willingness of even the best-intentioned taxpayer to figure out what he or she should pay. . . .

Since 1981, Washington has put ten major tax laws on the books, generating changes whose cumulative effect "is pretty staggering for the small businessperson," said Edward Koos, a tax-policy lawyer at the Small Business Administration.

Harold Apolinsky of the Small Business Council says 9,371 code sections have been amended since 1981, a total that owes much to the lobbying and campaign contributions of the powerful. "It appears to me that small business just has no clout," Mr. Apolinsky said. "Big business tolerates it," he added, referring to the resulting complexity. "Small business really can't."

SOURCE: *The New York Times*, January 30, 1994, Business Section, p. 4.

For example, suppose that the government taxes 20 percent of the first \$50,000 of income and 50 percent of all income above \$50,000. Under this tax, a person who makes \$60,000 pays a tax of \$15,000. (The tax equals $0.20 \times \$50,000$ plus $0.50 \times \$10,000$.) For this person, the average tax rate is $\$15,000/\$60,000$, or 25 percent. But

the marginal tax rate is 50 percent because the amount of the tax would rise by \$0.50 if the taxpayer earned an additional dollar.

The marginal and average tax rates each contain a useful piece of information. If we are trying to gauge the sacrifice made by a taxpayer, the average tax rate is more appropriate because it measures the fraction of income paid in taxes. By contrast, if we are trying to gauge how much the tax system distorts incentives, the marginal tax rate is more meaningful. One of the *Ten Principles of Economics* in Chapter 1 is that rational people think at the margin. A corollary to this principle is that the marginal tax rate measures how much the tax system discourages people from working hard. It is the marginal tax rate, therefore, that determines the deadweight loss of an income tax.



lump-sum tax

a tax that is the same amount for every person

LUMP-SUM TAXES

Suppose the government imposes a tax of \$4,000 on everyone. That is, everyone owes the same amount, regardless of earnings or any actions that a person might take. Such a tax is called a **lump-sum tax**.

A lump-sum tax shows clearly the difference between average and marginal tax rates. For a taxpayer with income of \$20,000, the average tax rate of a \$4,000 lump-sum tax is 20 percent; for a taxpayer with income of \$40,000, the average tax rate is 10 percent. For both taxpayers, the marginal tax rate is zero because an additional dollar of income would not change the amount of tax owed.

A lump-sum tax is the most efficient tax possible. Because a person's decisions do not alter the amount owed, the tax does not distort incentives and, therefore, does not cause deadweight losses. Because everyone can easily compute the amount owed and because there is no benefit to hiring tax lawyers and accountants, the lump-sum tax imposes a minimal administrative burden on taxpayers.

If lump-sum taxes are so efficient, why do we rarely observe them in the real world? The reason is that efficiency is only one goal of the tax system. A lump-sum tax would take the same amount from the poor and the rich, an outcome most people would view as unfair. To understand the tax systems that we observe, we must therefore consider the other major goal of tax policy: equity.

QUICK QUIZ: What is meant by the *efficiency* of a tax system? ♦ What can make a tax system inefficient?

TAXES AND EQUITY

Ever since American colonists dumped imported tea into Boston harbor to protest high British taxes, tax policy has generated some of the most heated debates in American politics. The heat is rarely fueled by questions of efficiency. Instead, it arises from disagreements over how the tax burden should be distributed. Senator Russell Long once mimicked the public debate with this ditty:

Don't tax you.
Don't tax me.
Tax that fella behind the tree.

Of course, if we are to rely on the government to provide some of the goods and services we want, taxes must fall on someone. In this section we consider the equity of a tax system. How should the burden of taxes be divided among the population? How do we evaluate whether a tax system is fair? Everyone agrees that the tax system should be equitable, but there is much disagreement about what equity means and how the equity of a tax system can be judged.

THE BENEFITS PRINCIPLE

One principle of taxation, called the **benefits principle**, states that people should pay taxes based on the benefits they receive from government services. This principle tries to make public goods similar to private goods. It seems fair that a person who often goes to the movies pays more in total for movie tickets than a person who rarely goes. Similarly, a person who gets great benefit from a public good should pay more for it than a person who gets little benefit.

The gasoline tax, for instance, is sometimes justified using the benefits principle. In some states, revenues from the gasoline tax are used to build and maintain roads. Because those who buy gasoline are the same people who use the roads, the gasoline tax might be viewed as a fair way to pay for this government service.

The benefits principle can also be used to argue that wealthy citizens should pay higher taxes than poorer ones. Why? Simply because the wealthy benefit more from public services. Consider, for example, the benefits of police protection from theft. Citizens with much to protect get greater benefit from police than do those with less to protect. Therefore, according to the benefits principle, the wealthy should contribute more than the poor to the cost of maintaining the police force. The same argument can be used for many other public services, such as fire protection, national defense, and the court system.

It is even possible to use the benefits principle to argue for antipoverty programs funded by taxes on the wealthy. As we discussed in Chapter 11, people prefer living in a society without poverty, suggesting that antipoverty programs are a public good. If the wealthy place a greater dollar value on this public good than members of the middle class do, perhaps just because the wealthy have more to spend, then, according to the benefits principle, they should be taxed more heavily to pay for these programs.

benefits principle

the idea that people should pay taxes based on the benefits they receive from government services

THE ABILITY-TO-PAY PRINCIPLE

Another way to evaluate the equity of a tax system is called the **ability-to-pay principle**, which states that taxes should be levied on a person according to how well that person can shoulder the burden. This principle is sometimes justified by the claim that all citizens should make an “equal sacrifice” to support the government. The magnitude of a person's sacrifice, however, depends not only on the size of his tax payment but also on his income and other circumstances. A \$1,000 tax paid by a poor person may require a larger sacrifice than a \$10,000 tax paid by a rich one.

ability-to-pay principle

the idea that taxes should be levied on a person according to how well that person can shoulder the burden

INCOME	PROPORTIONAL TAX		REGRESSIVE TAX		PROGRESSIVE TAX	
	AMOUNT OF TAX	PERCENT OF INCOME	AMOUNT OF TAX	PERCENT OF INCOME	AMOUNT OF TAX	PERCENT OF INCOME
\$ 50,000	\$12,500	25%	\$15,000	30%	\$10,000	20%
100,000	25,000	25	25,000	25	25,000	25
200,000	50,000	25	40,000	20	60,000	30

Table 12-7

THREE TAX SYSTEMS

vertical equity

the idea that taxpayers with a greater ability to pay taxes should pay larger amounts

horizontal equity

the idea that taxpayers with similar abilities to pay taxes should pay the same amount

proportional tax

a tax for which high-income and low-income taxpayers pay the same fraction of income

regressive tax

a tax for which high-income taxpayers pay a smaller fraction of their income than do low-income taxpayers

progressive tax

a tax for which high-income taxpayers pay a larger fraction of their income than do low-income taxpayers

The ability-to-pay principle leads to two corollary notions of equity: vertical equity and horizontal equity. **Vertical equity** states that taxpayers with a greater ability to pay taxes should contribute a larger amount. **Horizontal equity** states that taxpayers with similar abilities to pay should contribute the same amount. Although these notions of equity are widely accepted, applying them to evaluate a tax system is rarely straightforward.

Vertical Equity If taxes are based on ability to pay, then richer taxpayers should pay more than poorer taxpayers. But how much more should the rich pay? Much of the debate over tax policy concerns this question.

Consider the three tax systems in Table 12-7. In each case, taxpayers with higher incomes pay more. Yet the systems differ in how quickly taxes rise with income. The first system is called **proportional** because all taxpayers pay the same fraction of income. The second system is called **regressive** because high-income taxpayers pay a smaller fraction of their income, even though they pay a larger amount. The third system is called **progressive** because high-income taxpayers pay a larger fraction of their income.

Which of these three tax systems is most fair? There is no obvious answer, and economic theory does not offer any help in trying to find one. Equity, like beauty, is in the eye of the beholder.

CASE STUDY HOW THE TAX BURDEN IS DISTRIBUTED

Much debate over tax policy concerns whether the wealthy pay their fair share. There is no objective way to make this judgment. In evaluating the issue for yourself, however, it is useful to know how much families of different incomes pay under the current tax system.

Table 12-8 shows how all federal taxes are distributed among income classes. To construct this table, families are ranked according to their income and placed into five groups of equal size, called *quintiles*. The second column of the table shows the average income of each group. The poorest one-fifth of families had average income of \$9,880; the richest one-fifth had average income of \$174,000.

The next column of the table shows total taxes as a percent of income. As you can see, the U.S. federal tax system is progressive. The poorest families paid

Table 12-8

THE BURDEN OF FEDERAL TAXES

QUINTILE	AVERAGE INCOME	TAXES AS A PERCENT OF INCOME	PERCENT OF ALL INCOME	PERCENT OF ALL TAXES
Lowest	\$ 9,880	8.0%	4%	1%
Second	26,100	15.6	11	7
Middle	44,300	20.3	16	13
Fourth	68,200	23.1	20	19
Highest	174,000	29.1	49	59

SOURCE: Congressional Budget Office; estimates are for 1999.

8.0 percent of their incomes in taxes, and the richest paid 29.1 percent of their incomes.

The fourth and fifth columns compare the distribution of income and the distribution of taxes among these five groups. The poorest group earns 4 percent of all income and pays 1 percent of all taxes. The richest group earns 49 percent of all income and pays 59 percent of all taxes.

This table on taxes is a good starting point for understanding the burden of government, but the picture it offers is incomplete. Although it includes all the taxes that flow from households to the federal government, it fails to include the transfer payments, such as Social Security and welfare, that flow from the federal government back to households. Studies that include both taxes and transfers show more progressivity. The richest group of families still pays about one-quarter of its income to the government, even after transfers are subtracted. By contrast, poor families typically receive more in transfers than they pay in taxes. The average tax rate of the poorest quintile, rather than being 8.0 percent as in the table, is a *negative* 30 percent. In other words, their income is about 30 percent higher than it would be without government taxes and transfers. The lesson is clear: To understand fully the progressivity of government policies, one must take account of both what people pay and what they receive.

Horizontal Equity If taxes are based on ability to pay, then similar taxpayers should pay similar amounts of taxes. But what determines if two taxpayers are similar? Families differ in many ways. To evaluate whether a tax code is horizontally equitable, one must determine which differences are relevant for a family's ability to pay and which differences are not.

Suppose the Smith and Jones families each have income of \$50,000. The Smiths have no children, but Mr. Smith has an illness that causes medical expenses of \$20,000. The Joneses are in good health, but they have four children. Two of the Jones children are in college, generating tuition bills of \$30,000. Would it be fair for these two families to pay the same tax because they have the same income? Would it be more fair to give the Smiths a tax break to help them offset their high medical expenses? Would it be more fair to give the Joneses a tax break to help them with their tuition expenses?

There are no easy answers to these questions. In practice, the U.S. income tax is filled with special provisions that alter a family's tax based on its specific circumstances.

CASE STUDY HORIZONTAL EQUITY AND THE MARRIAGE TAX

The treatment of marriage provides an important example of how difficult it is to achieve horizontal equity in practice. Consider two couples who are exactly the same except that one couple is married and the other couple is not. A peculiar feature of the U.S. income tax code is that these two couples pay different taxes. The reason that marriage affects the tax liability of a couple is that the tax law treats a married couple as a single taxpayer. When a man and woman get married, they stop paying taxes as individuals and start paying taxes as a family. If the man and woman have similar incomes, their total tax liability rises when they get married.

To see how this “marriage tax” works, consider the following example of a progressive income tax. Suppose that the government taxes 25 percent of all income above \$10,000. Income below \$10,000 is excluded from taxation. Let’s see how this system treats two different couples.

Consider first Sam and Sally. Sam is a struggling poet and earns no income, whereas Sally is a lawyer and earns \$100,000 a year. Before getting married, Sam pays no tax. Sally pays 25 percent of \$90,000 (\$100,000 minus the \$10,000 exclusion), which is \$22,500. After getting married, their tax bill is the same. In this case, the income tax neither encourages nor discourages marriage.

Now consider John and Joan, two college professors each earning \$50,000 a year. Before getting married, they each pay a tax of \$10,000 (25 percent of \$40,000), or a total of \$20,000. After getting married, they have a total income of \$100,000, and so they owe a tax of 25 percent of \$90,000, or \$22,500. Thus, when John and Joan get married, their tax bill rises by \$2,500. This increase is called the marriage tax.



“And do you promise to love, honor, and cherish each other, and to pay the United States government more in taxes as a married couple than you would have paid if you had just continued living together?”

We can fix the problem for John and Joan by raising the income exclusion from \$10,000 to \$20,000 for married couples. But this change would create another problem. In this case, Sam and Sally would pay a tax after getting married of only \$20,000, which is \$2,500 less than they paid when they were single. Eliminating the marriage tax for John and Joan would create a marriage subsidy for Sam and Sally.

In practice, the U.S. tax code is an uneasy compromise that includes a combination of marriage taxes and marriage subsidies. According to a study by the Congressional Budget Office, 42 percent of married couples pay a marriage tax, averaging 2.0 percent of their income, while 51 percent of married couples pay lower taxes by virtue of being wed, averaging 2.3 percent of their income. Whether a couple is better off (from a tax standpoint) being married or shackled up depends on how earnings are split between the two partners. If a man and woman have similar incomes (like John and Joan), their wedding will most likely raise their tax bill. But a marriage subsidy is likely if one partner earns much more than the other, and especially if only one of them has earnings (like Sam and Sally).

This problem has no simple solution. To see why, try designing an income tax with the following four properties:

- ◆ Two married couples with the same total income should pay the same tax.
- ◆ When two people get married, their total tax bill should not change.
- ◆ A person or family with no income should pay no taxes.
- ◆ High-income taxpayers should pay a higher fraction of their incomes than low-income taxpayers.

All four of these properties are appealing, yet it is impossible to satisfy all of them simultaneously. Any income tax that satisfies the first three must violate the fourth. The only income tax that satisfies the first three properties is a proportional tax.

Some economists have advocated abolishing the marriage penalty by making individuals rather than the family the taxpaying unit, a policy that many European countries follow. This alternative might seem more equitable because it would treat married and unmarried couples the same. Yet this change would give up on the first of these properties: Families with the same total income could end up paying different taxes. For example, if each married couple paid taxes as if they were not married, then Sam and Sally would pay \$22,500, and John and Joan would pay \$20,000, even though both couples have the same total income. Whether this alternative tax system is more or less fair than the current marriage tax is hard to say.

TAX INCIDENCE AND TAX EQUITY

Tax incidence—the study of who bears the burden of taxes—is central to evaluating tax equity. As we first saw in Chapter 6, the person who bears the burden of a tax is not always the person who gets the tax bill from the government. Because taxes alter supply and demand, they alter equilibrium prices. As a result, they affect people beyond those who, according to statute, actually pay the tax. When

evaluating the vertical and horizontal equity of any tax, it is important to take account of these indirect effects.

Many discussions of tax equity ignore the indirect effects of taxes and are based on what economists mockingly call the *flypaper theory* of tax incidence. According to this theory, the burden of a tax, like a fly on flypaper, sticks wherever it first lands. This assumption, however, is rarely valid.

For example, a person not trained in economics might argue that a tax on expensive fur coats is vertically equitable because most buyers of furs are wealthy. Yet if these buyers can easily substitute other luxuries for furs, then a tax on furs might only reduce the sale of furs. In the end, the burden of the tax will fall more on those who make and sell furs than on those who buy them. Because most workers who make furs are not wealthy, the equity of a fur tax could be quite different from what the flypaper theory indicates.



THIS WORKER PAYS PART OF THE CORPORATE INCOME TAX.

CASE STUDY WHO PAYS THE CORPORATE INCOME TAX?

The corporate income tax provides a good example of the importance of tax incidence for tax policy. The corporate tax is popular among voters. After all, corporations are not people. Voters are always eager to have their taxes reduced and have some impersonal corporation pick up the tab.

But before deciding that the corporate income tax is a good way for the government to raise revenue, we should consider who bears the burden of the corporate tax. This is a difficult question on which economists disagree, but one thing is certain: *People pay all taxes*. When the government levies a tax on a corporation, the corporation is more like a tax collector than a taxpayer. The burden of the tax ultimately falls on people—the owners, customers, or workers of the corporation.

Many economists believe that workers and customers bear much of the burden of the corporate income tax. To see why, consider an example. Suppose that the U.S. government decides to raise the tax on the income earned by car companies. At first, this tax hurts the owners of the car companies, who receive less profit. But, over time, these owners will respond to the tax. Because producing cars is less profitable, they invest less in building new car factories. Instead, they invest their wealth in other ways—for example, by buying larger houses or by building factories in other industries or other countries. With fewer car factories, the supply of cars declines, as does the demand for autoworkers. Thus, a tax on corporations making cars causes the price of cars to rise and the wages of autoworkers to fall.

The corporate income tax shows how dangerous the flypaper theory of tax incidence can be. The corporate income tax is popular in part because it appears to be paid by rich corporations. Yet those who bear the ultimate burden of the tax—the customers and workers of corporations—are often not rich. If the true incidence of the corporate tax were more widely known, this tax might be less popular among voters.

CASE STUDY THE FLAT TAX

A recurring topic of debate is whether the U.S. federal government should completely scrap the current tax system and replace it with a much simpler system

called the *flat tax*. The flat tax was proposed in the early 1980s by economist Robert Hall and political scientist Alvin Rabushka. Since then, it has from time to time caught the attention of politicians on both the political left (such as Jerry Brown, former governor of California and sometime candidate in the Democratic presidential primaries) and the political right (such as Steve Forbes, multimillionaire publisher and sometime candidate in the Republican presidential primaries).

Although flat-tax advocates have proposed various plans that differ in detail, the essence of all the plans is a single, low tax rate that would apply to all income in the economy. If the tax rate were set at 19 percent, for example, then every taxpayer in the economy would face a marginal tax rate of 19 percent. Most of the plans allow a certain amount of income to be excluded from the tax. If the income exclusion were \$10,000, for instance, then a person's tax bill would be

$$\text{Tax} = 0.19 \times (\text{Income} - \$10,000).$$

Because of the income exclusion, a flat tax can be progressive: Average tax rates rise with income, even though the marginal tax rate is constant. Some of the plans even allow a person with very low income (in this example, less than \$10,000) to pay a "negative tax" by receiving a check from the government.

Because the flat-tax proposal calls for a major overhaul of the tax system, it raises almost every issue discussed in this chapter, especially the tradeoff between efficiency and equity. Here are some of the points made by flat-tax advocates:

- ◆ The flat tax would eliminate many of the deductions allowed under the current income tax, such as deductions for mortgage interest payments and charitable giving. By broadening the tax base in this way, the flat tax is able to reduce the marginal tax rates that most people face. Lower tax rates mean greater economic efficiency. Thus, flat-tax advocates claim that this change would expand the size of the economic pie.
- ◆ Because the flat tax is so simple, the administrative burden of taxation would be greatly reduced. Flat-tax advocates claim that many taxpayers could file their returns on a postcard. Because all taxpayers would pay the same low tax rate on all forms of income, people would have less incentive to hire tax lawyers and accountants to take advantage of loopholes.
- ◆ Because all taxpayers would face the same marginal tax rate, the tax could be collected at the source of income rather than from the person who receives the income. Income from corporate profit, for instance, would be taxed at the corporate level rather than at the personal level. This additional simplification also reduces administrative costs.
- ◆ The flat tax would replace both the personal income tax and the corporate income tax. All income, whether from working at a job or from owning shares in a corporation, would be taxed once at the same marginal rate. The flat tax would eliminate the current double taxation of corporate profits, which now discourages corporations from investing in new plants and equipment.

- ◆ In computing income for tax purposes, businesses would be allowed to deduct all legitimate business expenses, including expenses on new investment goods. This deduction for investment makes the flat tax more like a consumption tax than an income tax. As a result, a change to a flat tax would increase the incentive to save (or, more precisely, would eliminate the current tax system's disincentive to save).

In short, advocates of the flat tax claim that there is a strong efficiency argument for this dramatic tax reform.

Critics of the flat tax are sympathetic with the goal of a simpler and more efficient tax system, but they oppose the flat tax because they believe that it gives too little weight to the goal of vertical equity. They claim that a flat tax would be less progressive than the current tax system and, in particular, would shift some of the tax burden from the wealthy to the middle class. This concern may well be justified, but no one knows for sure. Our study of tax incidence shows that the burden of a tax is not necessarily borne by the person who sends the check to the government. If the flat tax did encourage greater saving, as advocates claim, it would lead to more rapid economic growth, which would benefit all taxpayers. No one can be certain, however, about how large the impact on economic growth would be.

QUICK QUIZ: Explain the *benefits principle* and the *ability-to-pay principle*.

- ◆ What are *vertical equity* and *horizontal equity*? ◆ Why is studying tax incidence important for determining the equity of a tax system?

CONCLUSION: THE TRADEOFF BETWEEN EQUITY AND EFFICIENCY

Almost everyone agrees that equity and efficiency are the two most important goals of the tax system. But often these two goals conflict. Many proposed changes in the tax laws increase efficiency while reducing equity, or increase equity while reducing efficiency. People disagree about tax policy often because they attach different weights to these two goals.

The recent history of tax policy shows how political leaders differ in their views on equity and efficiency. When Ronald Reagan was elected president in 1980, the marginal tax rate on the earnings of the richest Americans was 50 percent. On interest income, the marginal tax rate was 70 percent. Reagan argued that such high tax rates greatly distorted economic incentives to work and save. In other words, he claimed that these high tax rates cost too much in terms of economic efficiency. Tax reform was, therefore, a high priority of his administration. Reagan signed into law large cuts in tax rates in 1981 and then again in 1986. When Reagan left office in 1989, the richest Americans faced a marginal tax rate of only 28 percent. During the four years of the Bush presidency, the top tax rate increased slightly to 31 percent.

When Bill Clinton ran for president in 1992, he argued that the rich were not paying their fair share of taxes. In other words, the low tax rates on the rich

violated his view of vertical equity. One of President Clinton's first acts was to propose raising the tax rates on the highest levels of income. In 1993 the tax rates on the richest Americans became about 40 percent.

Economics alone cannot determine the best way to balance the goals of efficiency and equity. This issue involves political philosophy as well as economics. But economists do have an important role in the political debate over tax policy: They can shed light on the tradeoffs that society faces and can help us avoid policies that sacrifice efficiency without any benefit in terms of equity.

Summary

- ◆ The U.S. government raises revenue using various taxes. The most important taxes for the federal government are individual income taxes and payroll taxes for social insurance. The most important taxes for state and local governments are sales taxes and property taxes.
 - ◆ The efficiency of a tax system refers to the costs it imposes on taxpayers. There are two costs of taxes beyond the transfer of resources from the taxpayer to the government. The first is the distortion in the allocation of resources that arises as taxes alter incentives and behavior. The second is the administrative burden of complying with the tax laws.
 - ◆ The equity of a tax system concerns whether the tax burden is distributed fairly among the population.
- According to the benefits principle, it is fair for people to pay taxes based on the benefits they receive from the government. According to the ability-to-pay principle, it is fair for people to pay taxes based on their capability to handle the financial burden. When evaluating the equity of a tax system, it is important to remember a lesson from the study of tax incidence: The distribution of tax burdens is not the same as the distribution of tax bills.
- ◆ When considering changes in the tax laws, policymakers often face a tradeoff between efficiency and equity. Much of the debate over tax policy arises because people give different weights to these two goals.

Key Concepts

budget surplus, p. 248
 budget deficit, p. 248
 average tax rate, p. 252
 marginal tax rate, p. 252

lump-sum tax, p. 254
 benefits principle, p. 255
 ability-to-pay principle, p. 255
 vertical equity, p. 256

horizontal equity, p. 256
 proportional tax, p. 256
 regressive tax, p. 256
 progressive tax, p. 256

Questions for Review

1. Over the past several decades, has government grown more or less slowly than the rest of the economy?
2. What are the two most important sources of revenue for the U.S. federal government?
3. Explain how corporate profits are taxed twice.
4. Why is the burden of a tax to taxpayers greater than the revenue received by the government?
5. Why do some economists advocate taxing consumption rather than income?
6. Give two arguments why wealthy taxpayers should pay more taxes than poor taxpayers.
7. What is the concept of horizontal equity, and why is it hard to apply?
8. Describe the arguments for and against replacing the current tax system with a flat tax.

Problems and Applications

1. Government spending in the United States has grown as a share of national income over time. What changes in our economy and our society might explain this trend? Do you expect the trend to continue?
2. In a published source or on the Internet, find out whether the U.S. federal government had a budget deficit or surplus last year. What do policymakers expect to happen over the next few years? (Hint: The Web site of the Congressional Budget Office is www.cbo.gov.)
3. The information in many of the tables in this chapter is taken from the *Economic Report of the President*, which appears annually. Using a recent issue of the report at your library, answer the following questions and provide some numbers to support your answers.
 - a. Figure 12-1 shows that government revenue as a percentage of total income has increased over time. Is this increase primarily attributable to changes in federal government revenue or in state and local government revenue?
 - b. Looking at the combined revenue of the federal government and state and local governments, how has the composition of total revenue changed over time? Are personal income taxes more or less important? Social insurance taxes? Corporate profits taxes?
 - c. Looking at the combined expenditures of the federal government and state and local governments, how have the relative shares of transfer payments and purchases of goods and services changed over time?
4. The chapter states that the elderly population in the United States is growing more rapidly than the total population. In particular, the number of workers is rising slowly, while the number of retirees is rising quickly. Concerned about the future of Social Security, some members of Congress propose a “freeze” on the program.
 - a. If total expenditures were frozen, what would happen to benefits per retiree? To tax payments per worker? (Assume that Social Security taxes and receipts are balanced in each year.)
 - b. If benefits per retiree were frozen, what would happen to total expenditures? To tax payments per worker?
 - c. If tax payments per worker were frozen, what would happen to total expenditures? To benefits per retiree?
 - d. What do your answers to parts (a), (b), and (c) imply about the difficult decisions faced by policymakers?
5. Suppose you are a typical person in the U.S. economy. You pay a flat 4 percent of your income in a state income tax and 15.3 percent of your labor earnings in federal payroll taxes (employer and employee shares combined). You also pay federal income taxes as in Table 12-3. How much tax of each type do you pay if you earn \$20,000 a year? Taking all taxes into account, what are your average and marginal tax rates? What happens to your tax bill and to your average and marginal tax rates if your income rises to \$40,000?
6. Some states exclude necessities, such as food and clothing, from their sales tax. Other states do not. Discuss the merits of this exclusion. Consider both efficiency and equity.
7. Explain how individuals’ behavior is affected by the following features of the federal tax code.
 - a. Contributions to charity are tax deductible.
 - b. Sales of beer are taxed.
 - c. Interest that a homeowner pays on a mortgage is tax deductible.
 - d. Realized capital gains are taxed, but accrued gains are not. (When someone owns a share of stock that rises in value, she has an “accrued” capital gain. If she sells the share, she has a “realized” gain.)
8. Suppose that your state raises its sales tax from 5 percent to 6 percent. The state revenue commissioner forecasts a 20 percent increase in sales tax revenue. Is this plausible? Explain.
9. Consider two of the income security programs in the United States: Temporary Assistance for Needy Families (TANF) and the Earned Income Tax Credit (EITC).
 - a. When a woman with children and very low income earns an extra dollar, she receives less in TANF benefits. What do you think is the effect of this feature of TANF on the labor supply of low-income women? Explain.
 - b. The EITC provides greater benefits as low-income workers earn more income (up to a point). What do you think is the effect of this program on the labor supply of low-income individuals? Explain.

- c. What are the disadvantages of eliminating TANF and allocating the savings to the EITC?
10. The Tax Reform Act of 1986 eliminated the deductibility of interest payments on consumer debt (mostly credit cards and auto loans) but maintained the deductibility of interest payments on mortgages and home equity loans. What do you think happened to the relative amounts of borrowing through consumer debt and home equity debt?
11. Categorize each of the following funding schemes as examples of the benefits principle or the ability-to-pay principle.
 - a. Visitors to many national parks pay an entrance fee.
 - b. Local property taxes support elementary and secondary schools.
 - c. An airport trust fund collects a tax on each plane ticket sold and uses the money to improve airports and the air traffic control system.
12. Any income tax schedule embodies two types of tax rates—average tax rates and marginal tax rates.
 - a. The average tax rate is defined as total taxes paid divided by income. For the proportional tax system presented in Table 12-7, what are the average tax rates for people earning \$50,000, \$100,000, and \$200,000? What are the corresponding average tax rates in the regressive and progressive tax systems?
 - b. The marginal tax rate is defined as the extra taxes paid on additional income divided by the increase in income. Calculate the marginal tax rate for the proportional tax system as income rises from \$50,000 to \$100,000. Calculate the marginal tax rate as income rises from \$100,000 to \$200,000. Calculate the corresponding marginal tax rates for the regressive and progressive tax systems.
- c. Describe the relationship between average tax rates and marginal tax rates for each of these three systems. In general, which rate is relevant for someone deciding whether to accept a job that pays slightly more than her current job? Which rate is relevant for judging the vertical equity of a tax system?
13. What is the efficiency justification for taxing consumption rather than income? If the United States were to adopt a consumption tax, do you think that would make the U.S. tax system more or less progressive? Explain.
14. If a salesman takes a client to lunch, part of the cost of the lunch is a deductible business expense for his company. Some members of Congress have argued that this feature of the tax code benefits relatively wealthy businesspeople and should be eliminated. Yet their arguments have been met with greater opposition from eating and drinking establishments than from companies themselves. Explain.

13



THE COSTS OF PRODUCTION

The economy is made up of thousands of firms that produce the goods and services you enjoy every day: General Motors produces automobiles, General Electric produces lightbulbs, and General Mills produces breakfast cereals. Some firms, such as these three, are large; they employ thousands of workers and have thousands of stockholders who share in the firms' profits. Other firms, such as the local barbershop or candy store, are small; they employ only a few workers and are owned by a single person or family.

In previous chapters we used the supply curve to summarize firms' production decisions. According to the law of supply, firms are willing to produce and sell a greater quantity of a good when the price of the good is higher, and this response leads to a supply curve that slopes upward. For analyzing many questions, the law of supply is all you need to know about firm behavior.

In this chapter and the ones that follow, we examine firm behavior in more detail. This topic will give you a better understanding of what decisions lie behind

**IN THIS CHAPTER
YOU WILL . . .**

***Examine what items
are included in a
firm's costs of
production***

***Analyze the link
between a firm's
production process
and its total costs***

***Learn the meaning
of average total
cost and marginal
cost and how they
are related***

***Consider the shape
of a typical firm's
cost curves***

***Examine the
relationship
between short-run
and long-run costs***

the supply curve in a market. In addition, it will introduce you to a part of economics called *industrial organization*—the study of how firms' decisions regarding prices and quantities depend on the market conditions they face. The town in which you live, for instance, may have several pizzerias but only one cable television company. How does this difference in the number of firms affect the prices in these markets and the efficiency of the market outcomes? The field of industrial organization addresses exactly this question.

As a starting point for the study of industrial organization, this chapter examines the costs of production. All firms, from Delta Air Lines to your local deli, incur costs as they make the goods and services that they sell. As we will see in the coming chapters, a firm's costs are a key determinant of its production and pricing decisions. Establishing what a firm's costs are, however, is not as straightforward as it might seem.

WHAT ARE COSTS?

We begin our discussion of costs at Hungry Helen's Cookie Factory. Helen, the owner of the firm, buys flour, sugar, flavorings, and other cookie ingredients. She also buys the mixers and ovens and hires workers to run this equipment. She then sells the resulting cookies to consumers. By examining some of the issues that Helen faces in her business, we can learn some lessons that apply to all firms in the economy.

TOTAL REVENUE, TOTAL COST, AND PROFIT

We begin with the firm's objective. To understand what decisions a firm makes, we must understand what it is trying to do. It is conceivable that Helen started her firm because of an altruistic desire to provide the world with cookies or, perhaps, out of love for the cookie business. More likely, however, Helen started her business to make money. Economists normally assume that the goal of a firm is to maximize profit, and they find that this assumption works well in most cases.

What is a firm's profit? The amount that the firm receives for the sale of its output (cookies) is called its **total revenue**. The amount that the firm pays to buy inputs (flour, sugar, workers, ovens, etc.) is called its **total cost**. Helen gets to keep any revenue that is not needed to cover costs. We define **profit** as a firm's total revenue minus its total cost. That is,

$$\text{Profit} = \text{Total revenue} - \text{Total cost}.$$

Helen's objective is to make her firm's profit as large as possible.

To see how a firm goes about maximizing profit, we must consider fully how to measure its total revenue and its total cost. Total revenue is the easy part: It equals the quantity of output the firm produces times the price at which it sells its output. If Helen produces 10,000 cookies and sells them at \$2 a cookie, her total revenue is \$20,000. By contrast, the measurement of a firm's total cost is more subtle.

total revenue

the amount a firm receives for the sale of its output

total cost

the market value of the inputs a firm uses in production

profit

total revenue minus total cost

COSTS AS OPPORTUNITY COSTS

When measuring costs at Hungry Helen's Cookie Factory or any other firm, it is important to keep in mind one of the *Ten Principles of Economics* from Chapter 1: The cost of something is what you give up to get it. Recall that the *opportunity cost* of an item refers to all those things that must be forgone to acquire that item. When economists speak of a firm's cost of production, they include all the opportunity costs of making its output of goods and services.

A firm's opportunity costs of production are sometimes obvious and sometimes less so. When Helen pays \$1,000 for flour, that \$1,000 is an opportunity cost because Helen can no longer use that \$1,000 to buy something else. Similarly, when Helen hires workers to make the cookies, the wages she pays are part of the firm's costs. These are **explicit costs**. By contrast, some of a firm's opportunity costs are **implicit costs**. Imagine that Helen is skilled with computers and could earn \$100 per hour working as a programmer. For every hour that Helen works at her cookie factory, she gives up \$100 in income, and this forgone income is also part of her costs.

This distinction between explicit and implicit costs highlights an important difference between how economists and accountants analyze a business. Economists are interested in studying how firms make production and pricing decisions. Because these decisions are based on both explicit and implicit costs, economists include both when measuring a firm's costs. By contrast, accountants have the job of keeping track of the money that flows into and out of firms. As a result, they measure the explicit costs but often ignore the implicit costs.

The difference between economists and accountants is easy to see in the case of Hungry Helen's Cookie Factory. When Helen gives up the opportunity to earn money as a computer programmer, her accountant will not count this as a cost of her cookie business. Because no money flows out of the business to pay for this cost, it never shows up on the accountant's financial statements. An economist, however, will count the forgone income as a cost because it will affect the decisions that Helen makes in her cookie business. For example, if Helen's wage as a computer programmer rises from \$100 to \$500 per hour, she might decide that running her cookie business is too costly and choose to shut down the factory in order to become a full-time computer programmer.



explicit costs

input costs that require an outlay of money by the firm

implicit costs

input costs that do not require an outlay of money by the firm

THE COST OF CAPITAL AS AN OPPORTUNITY COST

An important implicit cost of almost every business is the opportunity cost of the financial capital that has been invested in the business. Suppose, for instance, that Helen used \$300,000 of her savings to buy her cookie factory from the previous owner. If Helen had instead left this money deposited in a savings account that pays an interest rate of 5 percent, she would have earned \$15,000 per year. To own her cookie factory, therefore, Helen has given up \$15,000 a year in interest income. This forgone \$15,000 is one of the implicit opportunity costs of Helen's business.

As we have already noted, economists and accountants treat costs differently, and this is especially true in their treatment of the cost of capital. An economist views the \$15,000 in interest income that Helen gives up every year as a cost of her business, even though it is an implicit cost. Helen's accountant, however, will not show this \$15,000 as a cost because no money flows out of the business to pay for it.

To further explore the difference between economists and accountants, let's change the example slightly. Suppose now that Helen did not have the entire

\$300,000 to buy the factory but, instead, used \$100,000 of her own savings and borrowed \$200,000 from a bank at an interest rate of 5 percent. Helen's accountant, who only measures explicit costs, will now count the \$10,000 interest paid on the bank loan every year as a cost because this amount of money now flows out of the firm. By contrast, according to an economist, the opportunity cost of owning the business is still \$15,000. The opportunity cost equals the interest on the bank loan (an explicit cost of \$10,000) plus the forgone interest on savings (an implicit cost of \$5,000).

ECONOMIC PROFIT VERSUS ACCOUNTING PROFIT

Now let's return to the firm's objective—profit. Because economists and accountants measure costs differently, they also measure profit differently. An economist measures a firm's **economic profit** as the firm's total revenue minus all the opportunity costs (explicit and implicit) of producing the goods and services sold. An accountant measures the firm's **accounting profit** as the firm's total revenue minus only the firm's explicit costs.

Figure 13-1 summarizes this difference. Notice that because the accountant ignores the implicit costs, accounting profit is larger than economic profit. For a business to be profitable from an economist's standpoint, total revenue must cover all the opportunity costs, both explicit and implicit.

economic profit

total revenue minus total cost, including both explicit and implicit costs

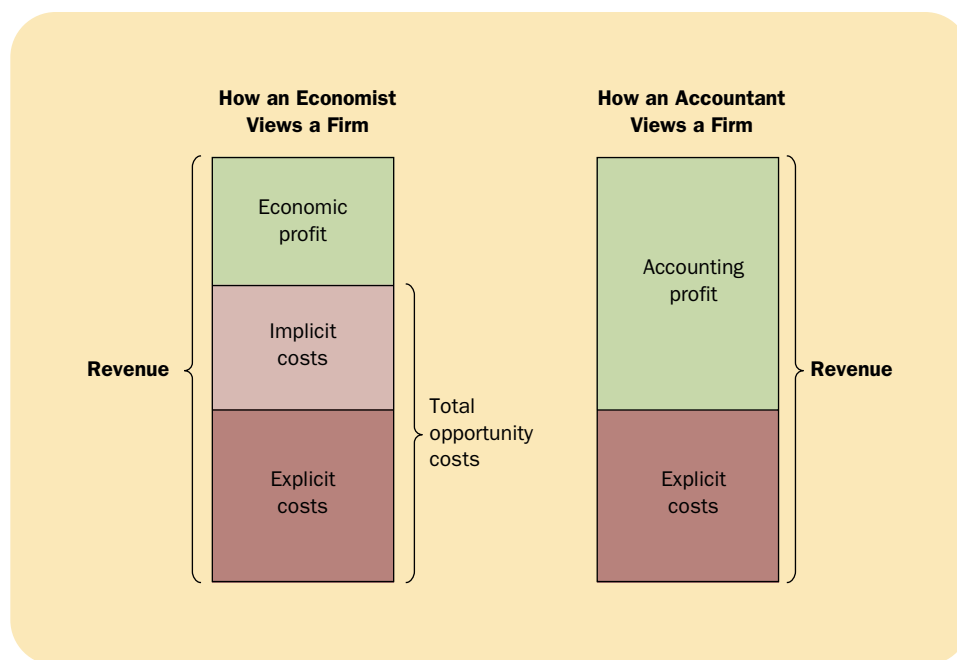
accounting profit

total revenue minus total explicit cost

QUICK QUIZ: Farmer McDonald gives banjo lessons for \$20 an hour. One day, he spends 10 hours planting \$100 worth of seeds on his farm. What opportunity cost has he incurred? What cost would his accountant measure? If these seeds will yield \$200 worth of crops, does McDonald earn an accounting profit? Does he earn an economic profit?

Figure 13-1

ECONOMISTS VERSUS ACCOUNTANTS. Economists include all opportunity costs when analyzing a firm, whereas accountants measure only explicit costs. Therefore, economic profit is smaller than accounting profit.



PRODUCTION AND COSTS

Firms incur costs when they buy inputs to produce the goods and services that they plan to sell. In this section we examine the link between a firm's production process and its total cost. Once again, we consider Hungry Helen's Cookie Factory.

In the analysis that follows, we make an important simplifying assumption: We assume that the size of Helen's factory is fixed and that Helen can vary the quantity of cookies produced only by changing the number of workers. This assumption is realistic in the short run, but not in the long run. That is, Helen cannot build a larger factory overnight, but she can do so within a year or so. This analysis, therefore, should be viewed as describing the production decisions that Helen faces in the short run. We examine the relationship between costs and time horizon more fully later in the chapter.

THE PRODUCTION FUNCTION

Table 13-1 shows how the quantity of cookies Helen's factory produces per hour depends on the number of workers. If there are no workers in the factory, Helen produces no cookies. When there is 1 worker, she produces 50 cookies. When there are 2 workers, she produces 90 cookies, and so on. Figure 13-2 presents a graph of these two columns of numbers. The number of workers is on the horizontal axis, and the number of cookies produced is on the vertical axis. This relationship between the quantity of inputs (workers) and quantity of output (cookies) is called the **production function**.

One of the *Ten Principles of Economics* introduced in Chapter 1 is that rational people think at the margin. As we will see in future chapters, this idea is the key to understanding the decision a firm makes about how many workers to hire and how much output to produce. To take a step toward understanding these decisions, the third column in the table gives the marginal product of a worker. The **marginal product** of any input in the production process is the increase in the quantity of output obtained from an additional unit of that input. When the number of workers goes from 1 to 2, cookie production increases from 50 to 90, so the marginal product of the second worker is 40 cookies. And when the number of workers goes from 2 to 3, cookie production increases from 90 to 120, so the marginal product of the third worker is 30 cookies.

Notice that as the number of workers increases, the marginal product declines. The second worker has a marginal product of 40 cookies, the third worker has a marginal product of 30 cookies, and the fourth worker has a marginal product of 20 cookies. This property is called **diminishing marginal product**. At first, when only a few workers are hired, they have easy access to Helen's kitchen equipment. As the number of workers increases, additional workers have to share equipment and work in more crowded conditions. Hence, as more and more workers are hired, each additional worker contributes less to the production of cookies.

Diminishing marginal product is also apparent in Figure 13-2. The production function's slope ("rise over run") tells us the change in Helen's output of

production function

the relationship between quantity of inputs used to make a good and the quantity of output of that good



marginal product

the increase in output that arises from an additional unit of input

diminishing marginal product

the property whereby the marginal product of an input declines as the quantity of the input increases

cookies (“rise”) for each additional input of labor (“run”). That is, the slope of the production function measures the marginal product of a worker. As the number of workers increases, the marginal product declines, and the production function becomes flatter.

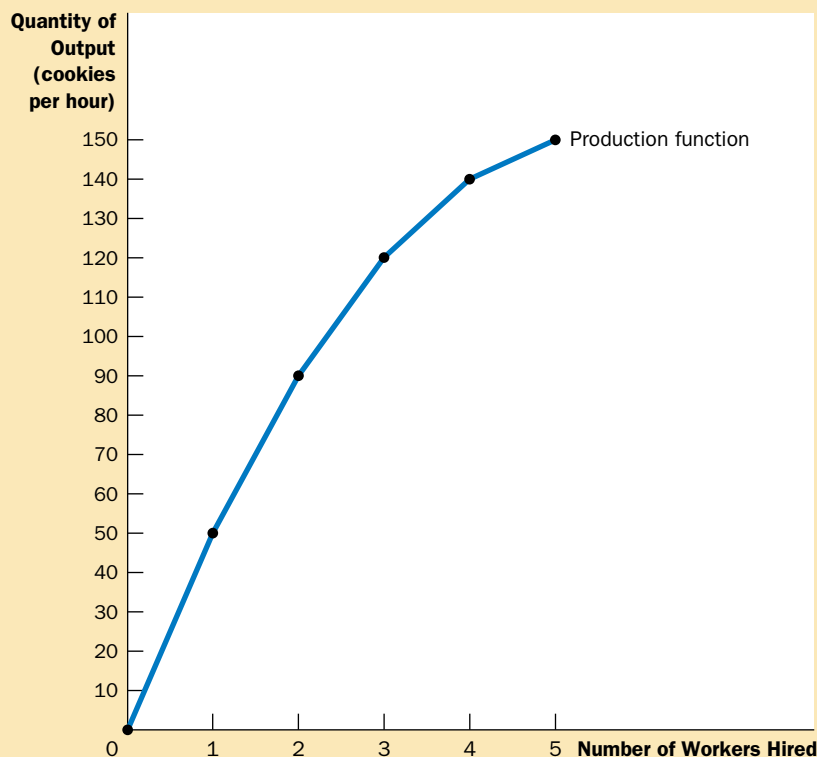
NUMBER OF WORKERS	OUTPUT (QUANTITY OF COOKIES PRODUCED PER HOUR)	MARGINAL PRODUCT OF LABOR	COST OF FACTORY	COST OF WORKERS	TOTAL COST OF INPUTS (COST OF FACTORY + COST OF WORKERS)
0	0		\$30	\$ 0	\$30
1	50	50	30	10	40
2	90	40	30	20	50
3	120	30	30	30	60
4	140	20	30	40	70
5	150	10	30	50	80

Table 13-1

A PRODUCTION FUNCTION AND TOTAL COST: HUNGRY HELEN’S COOKIE FACTORY

Figure 13-2

HUNGRY HELEN’S PRODUCTION FUNCTION. A production function shows the relationship between the number of workers hired and the quantity of output produced. Here the number of workers hired (on the horizontal axis) is from the first column in Table 13-1, and the quantity of output produced (on the vertical axis) is from the second column. The production function gets flatter as the number of workers increases, which reflects diminishing marginal product.



FROM THE PRODUCTION FUNCTION TO THE TOTAL-COST CURVE

The last three columns of Table 13-1 show Helen's cost of producing cookies. In this example, the cost of Helen's factory is \$30 per hour, and the cost of a worker is \$10 per hour. If she hires 1 worker, her total cost is \$40. If she hires 2 workers, her total cost is \$50, and so on. With this information, the table now shows how the number of workers Helen hires is related to the quantity of cookies she produces and to her total cost of production.

Our goal in the next several chapters is to study firms' production and pricing decisions. For this purpose, the most important relationship in Table 13-1 is between quantity produced (in the second column) and total costs (in the sixth column). Figure 13-3 graphs these two columns of data with the quantity produced on the horizontal axis and total cost on the vertical axis. This graph is called the *total-cost curve*.

Notice that the total cost gets steeper as the amount produced rises. The shape of the total-cost curve in this figure reflects the shape of the production function in Figure 13-2. Recall that when Helen's kitchen gets crowded, each additional worker adds less to the production of cookies; this property of diminishing marginal product is reflected in the flattening of the production function as the number of workers rises. But now turn this logic around: When Helen is producing a large quantity of cookies, she must have hired many workers. Because her kitchen is already crowded, producing an additional cookie is quite costly. Thus, as the quantity produced rises, the total-cost curve becomes steeper.

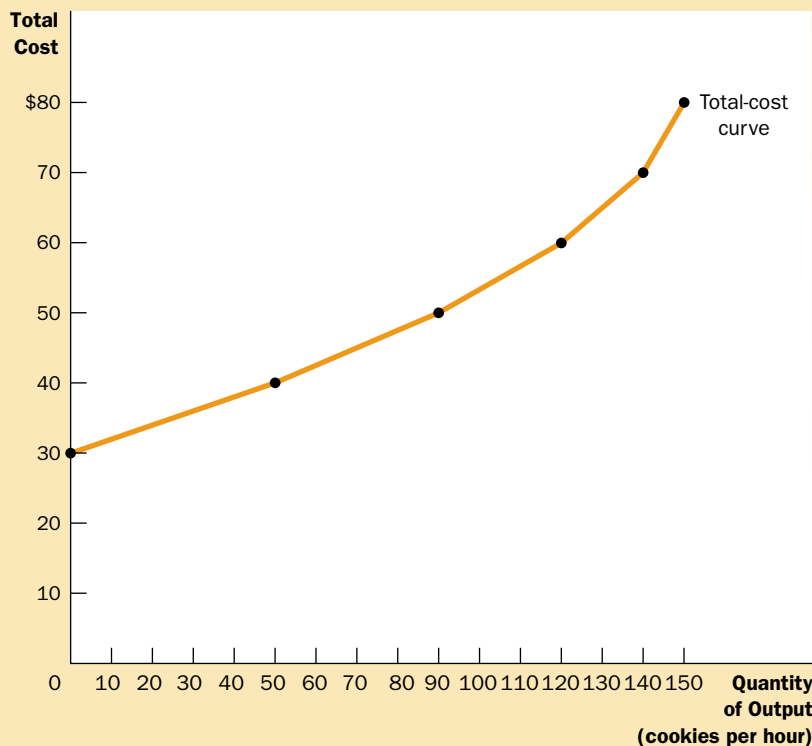


Figure 13-3

HUNGRY HELEN'S TOTAL-COST CURVE. A total-cost curve shows the relationship between the quantity of output produced and total cost of production. Here the quantity of output produced (on the horizontal axis) is from the second column in Table 13-1, and the total cost (on the vertical axis) is from the sixth column. The total-cost curve gets steeper as the quantity of output increases because of diminishing marginal product.

QUICK QUIZ: If Farmer Jones plants no seeds on his farm, he gets no harvest. If he plants 1 bag of seeds, he gets 3 bushels of wheat. If he plants 2 bags, he gets 5 bushels. If he plants 3 bags, he gets 6 bushels. A bag of seeds costs \$100, and seeds are his only cost. Use these data to graph the farmer's production function and total-cost curve. Explain their shapes.

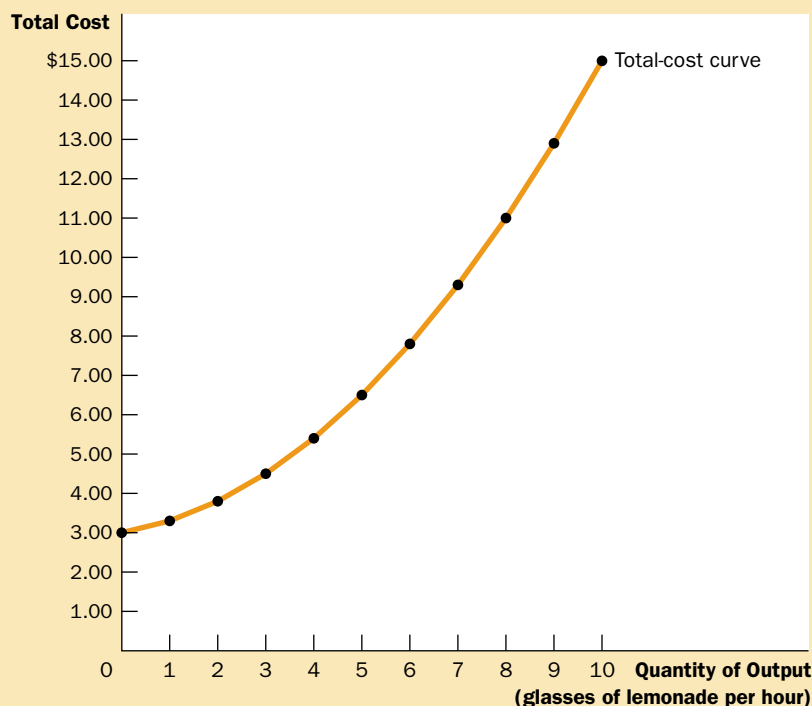
THE VARIOUS MEASURES OF COST

Our analysis of Hungry Helen's Cookie Factory demonstrated how a firm's total cost reflects its production function. From data on a firm's total cost, we can derive several related measures of cost, which will turn out to be useful when we analyze production and pricing decisions in future chapters. To see how these related measures are derived, we consider the example in Table 13-2. This table presents cost data on Helen's neighbor: Thirsty Thelma's Lemonade Stand.

The first column of the table shows the number of glasses of lemonade that Thelma might produce, ranging from 0 to 10 glasses per hour. The second column shows Thelma's total cost of producing lemonade. Figure 13-4 plots Thelma's total-cost curve. The quantity of lemonade (from the first column) is on the horizontal axis, and total cost (from the second column) is on the vertical axis. Thirsty

Figure 13-4

THIRSTY THELMA'S TOTAL-COST CURVE. Here the quantity of output produced (on the horizontal axis) is from the first column in Table 13-2, and the total cost (on the vertical axis) is from the second column. As in Figure 13-3, the total-cost curve gets steeper as the quantity of output increases because of diminishing marginal product.



QUANTITY OF LEMONADE (GLASSES PER HOUR)	TOTAL COST	FIXED COST	VARIABLE COST	AVERAGE FIXED COST	AVERAGE VARIABLE COST	AVERAGE TOTAL COST	MARGINAL COST
0	\$ 3.00	\$3.00	\$ 0.00	—	—	—	
1	3.30	3.00	0.30	\$3.00	\$0.30	\$3.30	\$0.30
2	3.80	3.00	0.80	1.50	0.40	1.90	0.50
3	4.50	3.00	1.50	1.00	0.50	1.50	0.70
4	5.40	3.00	2.40	0.75	0.60	1.35	0.90
5	6.50	3.00	3.50	0.60	0.70	1.30	1.10
6	7.80	3.00	4.80	0.50	0.80	1.30	1.30
7	9.30	3.00	6.30	0.43	0.90	1.33	1.50
8	11.00	3.00	8.00	0.38	1.00	1.38	1.70
9	12.90	3.00	9.90	0.33	1.10	1.43	1.90
10	15.00	3.00	12.00	0.30	1.20	1.50	2.10

THE VARIOUS MEASURES OF COST: THIRSTY THELMA'S LEMONADE STAND

Table 13-2

Thelma's total-cost curve has a shape similar to Hungry Helen's. In particular, it becomes steeper as the quantity produced rises, which (as we have discussed) reflects diminishing marginal product.

FIXED AND VARIABLE COSTS

Thelma's total cost can be divided into two types. Some costs, called **fixed costs**, do not vary with the quantity of output produced. They are incurred even if the firm produces nothing at all. Thelma's fixed costs include the rent she pays because this cost is the same regardless of how much lemonade Thelma produces. Similarly, if Thelma needs to hire a full-time bookkeeper to pay bills, regardless of the quantity of lemonade produced, the bookkeeper's salary is a fixed cost. The third column in Table 13-2 shows Thelma's fixed cost, which in this example is \$3.00 per hour.

Some of the firm's costs, called **variable costs**, change as the firm alters the quantity of output produced. Thelma's variable costs include the cost of lemons and sugar: The more lemonade Thelma makes, the more lemons and sugar she needs to buy. Similarly, if Thelma has to hire more workers to make more lemonade, the salaries of these workers are variable costs. The fourth column of the table shows Thelma's variable cost. The variable cost is 0 if she produces nothing, \$0.30 if she produces 1 glass of lemonade, \$0.80 if she produces 2 glasses, and so on.

A firm's total cost is the sum of fixed and variable costs. In Table 13-2, total cost in the second column equals fixed cost in the third column plus variable cost in the fourth column.

fixed costs

costs that do not vary with the quantity of output produced

variable costs

costs that do vary with the quantity of output produced

AVERAGE AND MARGINAL COST

As the owner of her firm, Thelma has to decide how much to produce. A key part of this decision is how her costs will vary as she changes the level of production. In making this decision, Thelma might ask her production supervisor the following two questions about the cost of producing lemonade:

- ◆ How much does it cost to make the typical glass of lemonade?
- ◆ How much does it cost to increase production of lemonade by 1 glass?

Although at first these two questions might seem to have the same answer, they do not. Both answers will turn out to be important for understanding how firms make production decisions.

To find the cost of the typical unit produced, we would divide the firm's costs by the quantity of output it produces. For example, if the firm produces 2 glasses per hour, its total cost is \$3.80, and the cost of the typical glass is \$3.80/2, or \$1.90. Total cost divided by the quantity of output is called **average total cost**. Because total cost is just the sum of fixed and variable costs, average total cost can be expressed as the sum of average fixed cost and average variable cost. **Average fixed cost** is the fixed cost divided by the quantity of output, and **average variable cost** is the variable cost divided by the quantity of output.

Although average total cost tells us the cost of the typical unit, it does not tell us how much total cost will change as the firm alters its level of production. The last column in Table 13-2 shows the amount that total cost rises when the firm increases production by 1 unit of output. This number is called **marginal cost**. For example, if Thelma increases production from 2 to 3 glasses, total cost rises from \$3.80 to \$4.50, so the marginal cost of the third glass of lemonade is \$4.50 minus \$3.80, or \$0.70.

It may be helpful to express these definitions mathematically. If Q stands for quantity, TC for total cost, ATC for average total cost, and MC for marginal cost, then we can then write:

$$ATC = \text{Total cost/Quantity} = TC/Q$$

and

$$MC = (\text{Change in total cost})/(\text{Change in quantity}) = \Delta TC/\Delta Q.$$

Here Δ , the Greek letter *delta*, represents the change in a variable. These equations show how average total cost and marginal cost are derived from total cost.

As we will see more fully in the next chapter, Thelma, our lemonade entrepreneur, will find the concepts of average total cost and marginal cost extremely useful when deciding how much lemonade to produce. Keep in mind, however, that these concepts do not actually give Thelma new information about her costs of production. Instead, average total cost and marginal cost express in a new way information that is already contained in her firm's total cost. *Average total cost tells us the cost of a typical unit of output if total cost is divided evenly over all the units produced. Marginal cost tells us the increase in total cost that arises from producing an additional unit of output.*

average total cost

total cost divided by the quantity of output

average fixed cost

fixed costs divided by the quantity of output

average variable cost

variable costs divided by the quantity of output

marginal cost

the increase in total cost that arises from an extra unit of production

COST CURVES AND THEIR SHAPES

Just as in previous chapters we found graphs of supply and demand useful when analyzing the behavior of markets, we will find graphs of average and marginal cost useful when analyzing the behavior of firms. Figure 13-5 graphs Thelma's costs using the data from Table 13-2. The horizontal axis measures the quantity the firm produces, and the vertical axis measures marginal and average costs. The graph shows four curves: average total cost (ATC), average fixed cost (AFC), average variable cost (AVC), and marginal cost (MC).

The cost curves shown here for Thirsty Thelma's Lemonade Stand have some features that are common to the cost curves of many firms in the economy. Let's examine three features in particular: the shape of marginal cost, the shape of average total cost, and the relationship between marginal and average total cost.

Rising Marginal Cost Thirsty Thelma's marginal cost rises with the quantity of output produced. This reflects the property of diminishing marginal product. When Thelma is producing a small quantity of lemonade, she has few workers, and much of her equipment is not being used. Because she can easily put these idle resources to use, the marginal product of an extra worker is large, and the marginal cost of an extra glass of lemonade is small. By contrast, when Thelma is producing a large quantity of lemonade, her stand is crowded with workers, and most of her equipment is fully utilized. Thelma can produce more lemonade by adding workers, but these new workers have to work in crowded conditions and may have to

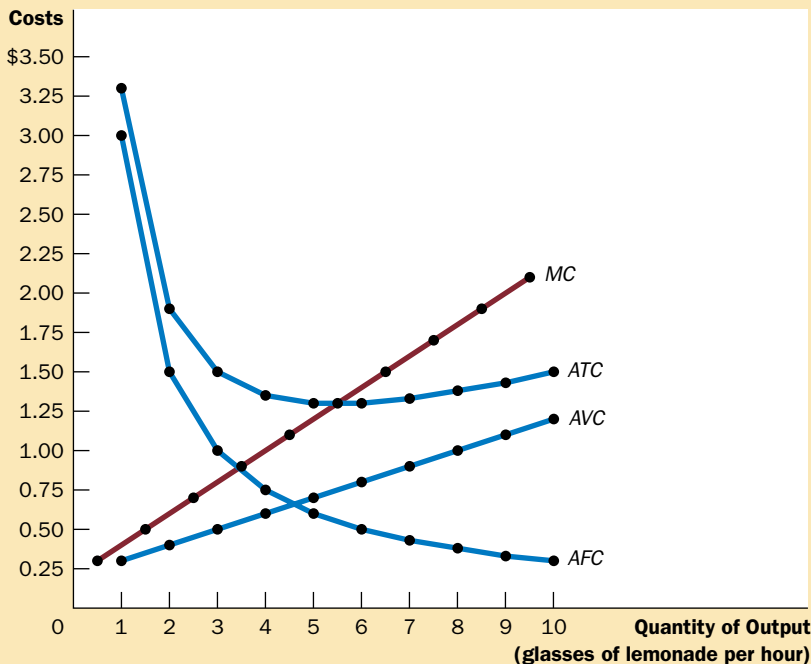


Figure 13-5

THIRSTY THELMA'S AVERAGE-COST AND MARGINAL-COST CURVES. This figure shows the average total cost (ATC), average fixed cost (AFC), average variable cost (AVC), and marginal cost (MC) for Thirsty Thelma's Lemonade Stand. All of these curves are obtained by graphing the data in Table 13-2. These cost curves show three features that are considered common: (1) Marginal cost rises with the quantity of output. (2) The average-total-cost curve is U-shaped. (3) The marginal-cost curve crosses the average-total-cost curve at the minimum of average total cost.

wait to use the equipment. Therefore, when the quantity of lemonade being produced is already high, the marginal product of an extra worker is low, and the marginal cost of an extra glass of lemonade is large.

U-Shaped Average Total Cost Thirsty Thelma's average-total-cost curve is U-shaped. To understand why this is so, remember that average total cost is the sum of average fixed cost and average variable cost. Average fixed cost always declines as output rises because the fixed cost is getting spread over a larger number of units. Average variable cost typically rises as output increases because of diminishing marginal product. Average total cost reflects the shapes of both average fixed cost and average variable cost. At very low levels of output, such as 1 or 2 glasses per hour, average total cost is high because the fixed cost is spread over only a few units. Average total cost then declines as output increases until the firm's output reaches 5 glasses of lemonade per hour, when average total cost falls to \$1.30 per glass. When the firm produces more than 6 glasses, average total cost starts rising again because average variable cost rises substantially.

The bottom of the U-shape occurs at the quantity that minimizes average total cost. This quantity is sometimes called the **efficient scale** of the firm. For Thirsty Thelma, the efficient scale is 5 or 6 glasses of lemonade. If she produces more or less than this amount, her average total cost rises above the minimum of \$1.30.

efficient scale

the quantity of output that minimizes average total cost

The Relationship between Marginal Cost and Average Total Cost If you look at Figure 13-5 (or back at Table 13-2), you will see something that may be surprising at first. *Whenever marginal cost is less than average total cost, average total cost is falling. Whenever marginal cost is greater than average total cost, average total cost is rising.* This feature of Thirsty Thelma's cost curves is not a coincidence from the particular numbers used in the example: It is true for all firms.

To see why, consider an analogy. Average total cost is like your cumulative grade point average. Marginal cost is like the grade in the next course you will take. If your grade in your next course is less than your grade point average, your grade point average will fall. If your grade in your next course is higher than your grade point average, your grade point average will rise. The mathematics of average and marginal costs is exactly the same as the mathematics of average and marginal grades.

This relationship between average total cost and marginal cost has an important corollary: *The marginal-cost curve crosses the average-total-cost curve at the efficient scale.* Why? At low levels of output, marginal cost is below average total cost, so average total cost is falling. But after the two curves cross, marginal cost rises above average total cost. For the reason we have just discussed, average total cost must start to rise at this level of output. Hence, this point of intersection is the minimum of average total cost. As you will see in the next chapter, this point of minimum average total cost plays a key role in the analysis of competitive firms.

TYPICAL COST CURVES

In the examples we have studied so far, the firms exhibit diminishing marginal product and, therefore, rising marginal cost at all levels of output. Yet actual firms are often a bit more complicated than this. In many firms, diminishing marginal product does not start to occur immediately after the first worker is hired. Depending on the

production process, the second or third worker might have higher marginal product than the first because a team of workers can divide tasks and work more productively than a single worker. Such firms would first experience increasing marginal product for a while before diminishing marginal product sets in.

Table 13-3 shows the cost data for such a firm, called Big Bob's Bagel Bin. These data are graphed in Figure 13-6. Panel (a) shows how total cost (TC) depends on the quantity produced, and panel (b) shows average total cost (ATC), average fixed cost (AFC), average variable cost (AVC), and marginal cost (MC). In the range of output from 0 to 4 bagels per hour, the firm experiences increasing marginal product, and the marginal-cost curve falls. After 5 bagels per hour, the firm starts to experience diminishing marginal product, and the marginal-cost curve starts to rise. This combination of increasing then diminishing marginal product also makes the average-variable-cost curve U-shaped.

Despite these differences from our previous example, Big Bob's cost curves share the three properties that are most important to remember:

- ◆ Marginal cost eventually rises with the quantity of output.
- ◆ The average-total-cost curve is U-shaped.
- ◆ The marginal-cost curve crosses the average-total-cost curve at the minimum of average total cost.

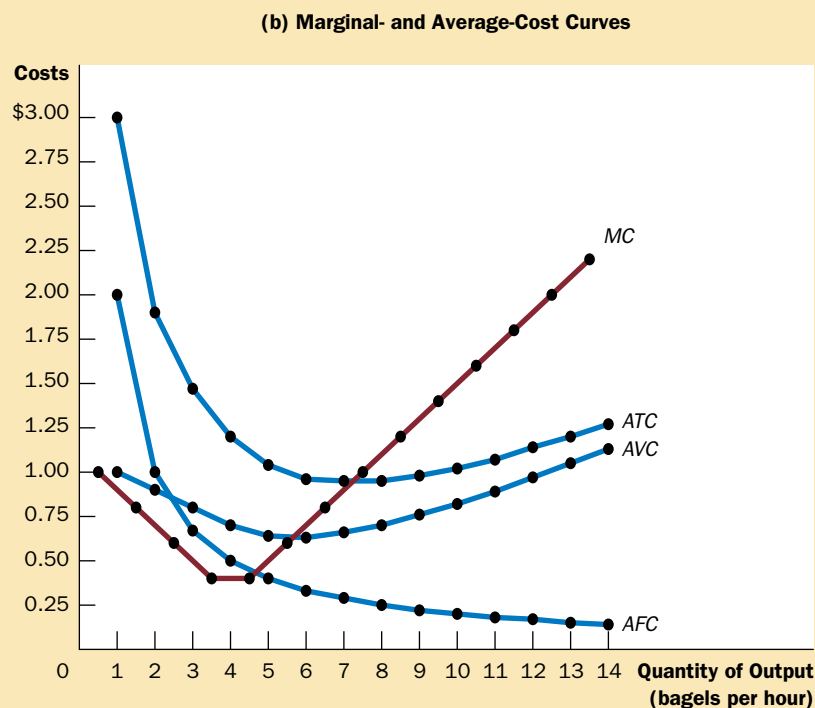
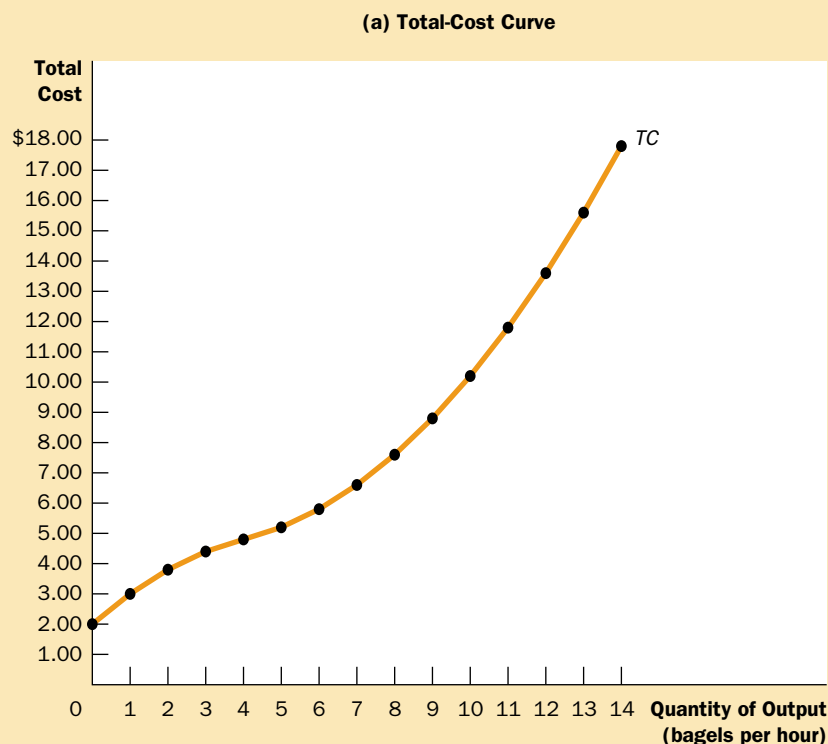
QUANTITY OF BAGELS (PER HOUR)	TOTAL COST	FIXED COST	VARIABLE COST	AVERAGE FIXED COST	AVERAGE VARIABLE COST	AVERAGE TOTAL COST	MARGINAL COST
0	\$ 2.00	\$2.00	\$ 0.00	—	—	—	\$1.00
1	3.00	2.00	1.00	\$2.00	\$1.00	\$3.00	0.80
2	3.80	2.00	1.80	1.00	0.90	1.90	0.60
3	4.40	2.00	2.40	0.67	0.80	1.47	0.40
4	4.80	2.00	2.80	0.50	0.70	1.20	0.40
5	5.20	2.00	3.20	0.40	0.64	1.04	0.60
6	5.80	2.00	3.80	0.33	0.63	0.96	0.80
7	6.60	2.00	4.60	0.29	0.66	0.95	1.00
8	7.60	2.00	5.60	0.25	0.70	0.95	1.20
9	8.80	2.00	6.80	0.22	0.76	0.98	1.40
10	10.20	2.00	8.20	0.20	0.82	1.02	1.60
11	11.80	2.00	9.80	0.18	0.89	1.07	1.80
12	13.60	2.00	11.60	0.17	0.97	1.14	2.00
13	15.60	2.00	13.60	0.15	1.05	1.20	2.20
14	17.80	2.00	15.80	0.14	1.13	1.27	

THE VARIOUS MEASURES OF COST: BIG BOB'S BAGEL BIN

Table 13-3

Figure 13-6

BIG BOB'S COST CURVES. Many firms, like Big Bob's Bagel Bin, experience increasing marginal product before diminishing marginal product and, therefore, have cost curves like those in this figure. Panel (a) shows how total cost (TC) depends on the quantity produced. Panel (b) shows how average total cost (ATC), average fixed cost (AFC), average variable cost (AVC), and marginal cost (MC) depend on the quantity produced. These curves are derived by graphing the data from Table 13-3. Notice that marginal cost and average variable cost fall for a while before starting to rise.



QUICK QUIZ: Suppose Honda's total cost of producing 4 cars is \$225,000 and its total cost of producing 5 cars is \$250,000. What is the average total cost of producing 5 cars? What is the marginal cost of the fifth car? ♦ Draw the marginal-cost curve and the average-total-cost curve for a typical firm, and explain why these curves cross where they do.

COSTS IN THE SHORT RUN AND IN THE LONG RUN

We noted at the beginning of this chapter that a firm's costs might depend on the time horizon being examined. Let's examine more precisely why this might be the case.

THE RELATIONSHIP BETWEEN SHORT-RUN AND LONG-RUN AVERAGE TOTAL COST

For many firms, the division of total costs between fixed and variable costs depends on the time horizon. Consider, for instance, a car manufacturer, such as Ford Motor Company. Over a period of only a few months, Ford cannot adjust the number or sizes of its car factories. The only way it can produce additional cars is to hire more workers at the factories it already has. The cost of these factories is, therefore, a fixed cost in the short run. By contrast, over a period of several years, Ford can expand the size of its factories, build new factories, or close old ones. Thus, the cost of its factories is a variable cost in the long run.

Because many decisions are fixed in the short run but variable in the long run, a firm's long-run cost curves differ from its short-run cost curves. Figure 13-7 shows an example. The figure presents three short-run average-total-cost curves—for a small, medium, and large factory. It also presents the long-run average-total-cost curve. As the firm moves along the long-run curve, it is adjusting the size of the factory to the quantity of production.

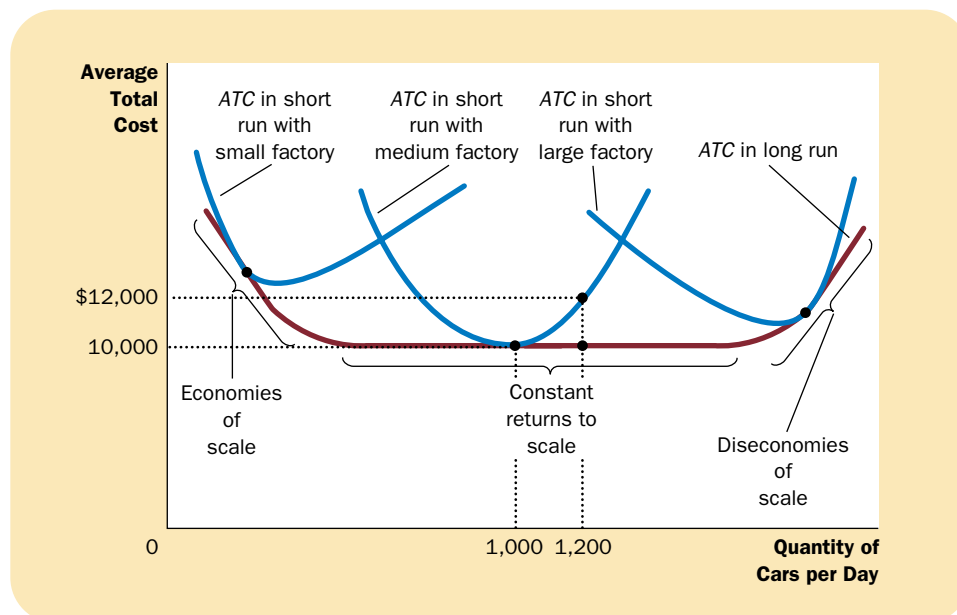
This graph shows how short-run and long-run costs are related. The long-run average-total-cost curve is a much flatter U-shape than the short-run average-total-cost curve. In addition, all the short-run curves lie on or above the long-run curve. These properties arise because of the greater flexibility firms have in the long run. In essence, in the long run, the firm gets to choose which short-run curve it wants to use. But in the short run, it has to use whatever short-run curve it chose in the past.

The figure shows an example of how a change in production alters costs over different time horizons. When Ford wants to increase production from 1,000 to 1,200 cars per day, it has no choice in the short run but to hire more workers at its existing medium-sized factory. Because of diminishing marginal product, average total cost rises from \$10,000 to \$12,000 per car. In the long run, however, Ford can expand both the size of the factory and its workforce, and average total cost remains at \$10,000.

How long does it take for a firm to get to the long run? The answer depends on the firm. It can take a year or longer for a major manufacturing firm, such as a

Figure 13-7

AVERAGE TOTAL COST IN THE SHORT AND LONG RUNS. Because fixed costs are variable in the long run, the average-total-cost curve in the short run differs from the average-total-cost curve in the long run.



car company, to build a larger factory. By contrast, a person running a lemonade stand can go and buy a larger pitcher within an hour or less. There is, therefore, no single answer about how long it takes a firm to adjust its production facilities.

ECONOMIES AND DISECONOMIES OF SCALE

economies of scale

the property whereby long-run average total cost falls as the quantity of output increases

diseconomies of scale

the property whereby long-run average total cost rises as the quantity of output increases

constant returns to scale

the property whereby long-run average total cost stays the same as the quantity of output changes

The shape of the long-run average-total-cost curve conveys important information about the technology for producing a good. When long-run average total cost declines as output increases, there are said to be **economies of scale**. When long-run average total cost rises as output increases, there are said to be **diseconomies of scale**. When long-run average total cost does not vary with the level of output, there are said to be **constant returns to scale**. In this example, Ford has economies of scale at low levels of output, constant returns to scale at intermediate levels of output, and diseconomies of scale at high levels of output.

What might cause economies or diseconomies of scale? Economies of scale often arise because higher production levels allow *specialization* among workers, which permits each worker to become better at his or her assigned tasks. For instance, modern assembly-line production requires a large number of workers. If Ford were producing only a small quantity of cars, it could not take advantage of this approach and would have higher average total cost. Diseconomies of scale can arise because of *coordination problems* that are inherent in any large organization. The more cars Ford produces, the more stretched the management team becomes, and the less effective the managers become at keeping costs down.

This analysis shows why long-run average-total-cost curves are often U-shaped. At low levels of production, the firm benefits from increased size because it can take advantage of greater specialization. Coordination problems,

FYI*Lessons from a
Pin Factory*

“Jack of all trades, master of none.” This well-known adage helps explain why firms sometimes experience economies of scale. A person who tries to do everything usually ends up doing nothing very well. If a firm wants its workers to be as productive as they can be, it is often best to give them a limited task that they can master. But this is possible only if a firm employs a large number of workers and

produces a large quantity of output.

In his celebrated book, *An Inquiry into the Nature and Causes of the Wealth of Nations*, Adam Smith described an example of this based on a visit he made to a pin factory. Smith was impressed by the specialization among the workers that he observed and the resulting economies of scale. He wrote,

“One man draws out the wire, another straightens it, a third cuts it, a fourth points it, a fifth grinds it at the top

for receiving the head; to make the head requires two or three distinct operations; to put it on is a peculiar business; to whiten it is another; it is even a trade by itself to put them into paper.”

Smith reported that because of this specialization, the pin factory produced thousands of pins per worker every day. He conjectured that if the workers had chosen to work separately, rather than as a team of specialists, “they certainly could not each of them make twenty, perhaps not one pin a day.” In other words, because of specialization, a large pin factory could achieve higher output per worker and lower average cost per pin than a small pin factory.

The specialization that Smith observed in the pin factory is prevalent in the modern economy. If you want to build a house, for instance, you could try to do all the work yourself. But most people turn to a builder, who in turn hires carpenters, plumbers, electricians, painters, and many other types of worker. These workers specialize in particular jobs, and this allows them to become better at their jobs than if they were generalists. Indeed, the use of specialization to achieve economies of scale is one reason modern societies are as prosperous as they are.

meanwhile, are not yet acute. By contrast, at high levels of production, the benefits of specialization have already been realized, and coordination problems become more severe as the firm grows larger. Thus, long-run average total cost is falling at low levels of production because of increasing specialization and rising at high levels of production because of increasing coordination problems.

QUICK QUIZ: If Boeing produces 9 jets per month, its long-run total cost is \$9.0 million per month. If it produces 10 jets per month, its long-run total cost is \$9.5 million per month. Does Boeing exhibit economies or diseconomies of scale?

CONCLUSION

The purpose of this chapter has been to develop some tools that we can use to study how firms make production and pricing decisions. You should now understand what economists mean by the term *costs* and how costs vary with the quantity of output a firm produces. To refresh your memory, Table 13-4 summarizes some of the definitions we have encountered.

Table 13-4

THE MANY TYPES OF COST:
A SUMMARY

TERM	DEFINITION	MATHEMATICAL DESCRIPTION
Explicit costs	Costs that require an outlay of money by the firm	—
Implicit costs	Costs that do not require an outlay of money by the firm	—
Fixed costs	Costs that do not vary with the quantity of output produced	FC
Variable costs	Costs that do vary with the quantity of output produced	VC
Total cost	The market value of all the inputs that a firm uses in production	$TC = FC + VC$
Average fixed cost	Fixed costs divided by the quantity of output	$AFC = FC/Q$
Average variable cost	Variable costs divided by the quantity of output	$AVC = VC/Q$
Average total cost	Total cost divided by the quantity of output	$ATC = TC/Q$
Marginal cost	The increase in total cost that arises from an extra unit of production	$MC = \Delta TC / \Delta Q$

By themselves, of course, a firm's cost curves do not tell us what decisions the firm will make. But they are an important component of that decision, as we will begin to see in the next chapter.

Summary

- ◆ The goal of firms is to maximize profit, which equals total revenue minus total cost.
- ◆ When analyzing a firm's behavior, it is important to include all the opportunity costs of production. Some of the opportunity costs, such as the wages a firm pays its workers, are explicit. Other opportunity costs, such as the wages the firm owner gives up by working in the firm rather than taking another job, are implicit.
- ◆ A firm's costs reflect its production process. A typical firm's production function gets flatter as the quantity of an input increases, displaying the property of diminishing marginal product. As a result, a firm's total-cost curve gets steeper as the quantity produced rises.
- ◆ A firm's total costs can be divided between fixed costs and variable costs. Fixed costs are costs that do not change when the firm alters the quantity of output produced. Variable costs are costs that do change when the firm alters the quantity of output produced.
- ◆ From a firm's total cost, two related measures of cost are derived. Average total cost is total cost divided by the quantity of output. Marginal cost is the amount by which total cost would rise if output were increased by 1 unit.
- ◆ When analyzing firm behavior, it is often useful to graph average total cost and marginal cost. For a typical firm, marginal cost rises with the quantity of output. Average total cost first falls as output increases and then

risers as output increases further. The marginal-cost curve always crosses the average-total-cost curve at the minimum of average total cost.

- ◆ A firm's costs often depend on the time horizon being considered. In particular, many costs are fixed in the

short run but variable in the long run. As a result, when the firm changes its level of production, average total cost may rise more in the short run than in the long run.

Key Concepts

total revenue, p. 270
total cost, p. 270
profit, p. 270
explicit costs, p. 271
implicit costs, p. 271
economic profit, p. 272
accounting profit, p. 272

production function, p. 273
marginal product, p. 273
diminishing marginal product, p. 273
fixed costs, p. 277
variable costs, p. 277
average total cost, p. 278
average fixed cost, p. 278

average variable cost, p. 278
marginal cost, p. 278
efficient scale, p. 280
economies of scale, p. 284
diseconomies of scale, p. 284
constant returns to scale, p. 284

Questions for Review

1. What is the relationship between a firm's total revenue, profit, and total cost?
2. Give an example of an opportunity cost that an accountant might not count as a cost. Why would the accountant ignore this cost?
3. What is marginal product, and what does it mean if it is diminishing?
4. Draw a production function that exhibits diminishing marginal product of labor. Draw the associated total-cost curve. (In both cases, be sure to label the axes.) Explain the shapes of the two curves you have drawn.
5. Define total cost, average total cost, and marginal cost. How are they related?
6. Draw the marginal-cost and average-total-cost curves for a typical firm. Explain why the curves have the shapes that they do and why they cross where they do.
7. How and why does a firm's average-total-cost curve differ in the short run and in the long run?
8. Define *economies of scale* and explain why they might arise. Define *diseconomies of scale* and explain why they might arise.

Problems and Applications

1. This chapter discusses many types of costs: opportunity cost, total cost, fixed cost, variable cost, average total cost, and marginal cost. Fill in the type of cost that best completes the phrases below:
 - a. The true cost of taking some action is its _____.
 - b. _____ is falling when marginal cost is below it, and rising when marginal cost is above it.
 - c. A cost that does not depend on the quantity produced is a _____.
 - d. In the ice-cream industry in the short run, _____ includes the cost of cream and sugar, but not the cost of the factory.
 - e. Profits equal total revenue less _____.
 - f. The cost of producing an extra unit of output is _____.
2. Your aunt is thinking about opening a hardware store. She estimates that it would cost \$500,000 per year to rent the location and buy the stock. In addition, she would have to quit her \$50,000 per year job as an accountant.
 - a. Define opportunity cost.
 - b. What is your aunt's opportunity cost of running a hardware store for a year? If your aunt thought she could sell \$510,000 worth of merchandise in a year, should she open the store? Explain.

3. Suppose that your college charges you separately for tuition and for room and board.
- What is a cost of attending college that is not an opportunity cost?
 - What is an explicit opportunity cost of attending college?
 - What is an implicit opportunity cost of attending college?
4. A commercial fisherman notices the following relationship between hours spent fishing and the quantity of fish caught:

HOURS	QUANTITY OF FISH (IN POUNDS)
0	0
1	10
2	18
3	24
4	28
5	30

- What is the marginal product of each hour spent fishing?
 - Use these data to graph the fisherman's production function. Explain its shape.
 - The fisherman has a fixed cost of \$10 (his pole). The opportunity cost of his time is \$5 per hour. Graph the fisherman's total-cost curve. Explain its shape.
5. Nimbus, Inc., makes brooms and then sells them door-to-door. Here is the relationship between the number of workers and Nimbus's output in a given day:

WORKERS	OUTPUT	MARGINAL PRODUCT	TOTAL COST	AVERAGE TOTAL COST	MARGINAL COST
0	0				
1	20				
2	50				
3	90				
4	120				
5	140				
6	150				
7	155				

- Fill in the column of marginal products. What pattern do you see? How might you explain it?
 - A worker costs \$100 a day, and the firm has fixed costs of \$200. Use this information to fill in the column for total cost.
 - Fill in the column for average total cost. (Recall that $ATC = TC/Q$.) What pattern do you see?
 - Now fill in the column for marginal cost. (Recall that $MC = \Delta TC / \Delta Q$.) What pattern do you see?
 - Compare the column for marginal product and the column for marginal cost. Explain the relationship.
 - Compare the column for average total cost and the column for marginal cost. Explain the relationship.
6. Suppose that you and your roommate have started a bagel delivery service on campus. List some of your fixed costs and describe why they are fixed. List some of your variable costs and describe why they are variable.
7. Consider the following cost information for a pizzeria:

Q (DOZENS)	TOTAL COST	VARIABLE COST
0	\$300	\$ 0
1	350	50
2	390	90
3	420	120
4	450	150
5	490	190
6	540	240

- What is the pizzeria's fixed cost?
 - Construct a table in which you calculate the marginal cost per dozen pizzas using the information on total cost. Also calculate the marginal cost per dozen pizzas using the information on variable cost. What is the relationship between these sets of numbers? Comment.
8. You are thinking about setting up a lemonade stand. The stand itself costs \$200. The ingredients for each cup of lemonade cost \$0.50.
- What is your fixed cost of doing business? What is your variable cost per cup?
 - Construct a table showing your total cost, average total cost, and marginal cost for output levels varying from zero to 10 gallons. (Hint: There are 16 cups in a gallon.) Draw the three cost curves.
9. Your cousin Vinnie owns a painting company with fixed costs of \$200 and the following schedule for variable costs:

	QUANTITY OF HOUSES PAINTED PER MONTH						
	1	2	3	4	5	6	7
Variable costs	\$10	\$20	\$40	\$80	\$160	\$320	\$640

Calculate average fixed cost, average variable cost, and average total cost for each quantity. What is the efficient scale of the painting company?

10. Healthy Harry's Juice Bar has the following cost schedules:

Q (VATS)	VARIABLE COST	TOTAL COST
0	\$ 0	\$ 30
1	10	40
2	25	55
3	45	75
4	70	100
5	100	130
6	135	165

- Calculate average variable cost, average total cost, and marginal cost for each quantity.
 - Graph all three curves. What is the relationship between the marginal-cost curve and the average-total-cost curve? Between the marginal-cost curve and the average-variable-cost curve? Explain.
11. Consider the following table of long-run total cost for three different firms:

	QUANTITY						
	1	2	3	4	5	6	7
Firm A	\$60	\$70	\$80	\$90	\$100	\$110	\$120
Firm B	11	24	39	56	75	96	119
Firm C	21	34	49	66	85	106	129

Does each of these firms experience economies of scale or diseconomies of scale?

14



FIRMS IN COMPETITIVE MARKETS

If your local gas station raised the price it charges for gasoline by 20 percent, it would see a large drop in the amount of gasoline it sold. Its customers would quickly switch to buying their gasoline at other gas stations. By contrast, if your local water company raised the price of water by 20 percent, it would see only a small decrease in the amount of water it sold. People might water their lawns less often and buy more water-efficient shower heads, but they would be hard-pressed to reduce water consumption greatly and would be unlikely to find another supplier. The difference between the gasoline market and the water market is obvious: There are many firms pumping gasoline, but there is only one firm pumping water. As you might expect, this difference in market structure shapes the pricing and production decisions of the firms that operate in these markets.

In this chapter we examine the behavior of competitive firms, such as your local gas station. You may recall that a market is competitive if each buyer and seller

**IN THIS CHAPTER
YOU WILL . . .**

***Learn what
characteristics
make a market
competitive***

***Examine how
competitive firms
decide how much
output to produce***

***Examine how
competitive firms
decide when to shut
down production
temporarily***

***Examine how
competitive firms
decide whether
to exit or enter
a market***

***See how firm
behavior determines
a market's short-
run and long-run
supply curves***

is small compared to the size of the market and, therefore, has little ability to influence market prices. By contrast, if a firm can influence the market price of the good it sells, it is said to have *market power*. In the three chapters that follow this one, we examine the behavior of firms with market power, such as your local water company.

Our analysis of competitive firms in this chapter will shed light on the decisions that lie behind the supply curve in a competitive market. Not surprisingly, we will find that a market supply curve is tightly linked to firms' costs of production. (Indeed, this general insight should be familiar to you from our analysis in Chapter 7.) But among a firm's various costs—fixed, variable, average, and marginal—which ones are most relevant for its decision about the quantity to supply? We will see that all these measures of cost play important and interrelated roles.

WHAT IS A COMPETITIVE MARKET?

Our goal in this chapter is to examine how firms make production decisions in competitive markets. As a background for this analysis, we begin by considering what a competitive market is.

THE MEANING OF COMPETITION

Although we have already discussed the meaning of competition in Chapter 4, let's review the lesson briefly. A **competitive market**, sometimes called a *perfectly competitive market*, has two characteristics:

- ◆ There are many buyers and many sellers in the market.
- ◆ The goods offered by the various sellers are largely the same.

As a result of these conditions, the actions of any single buyer or seller in the market have a negligible impact on the market price. Each buyer and seller takes the market price as given.

An example is the market for milk. No single buyer of milk can influence the price of milk because each buyer purchases a small amount relative to the size of the market. Similarly, each seller of milk has limited control over the price because many other sellers are offering milk that is essentially identical. Because each seller can sell all he wants at the going price, he has little reason to charge less, and if he charges more, buyers will go elsewhere. Buyers and sellers in competitive markets must accept the price the market determines and, therefore, are said to be *price takers*.

In addition to the foregoing two conditions for competition, there is a third condition sometimes thought to characterize perfectly competitive markets:

- ◆ Firms can freely enter or exit the market.

competitive market

a market with many buyers and sellers trading identical products so that each buyer and seller is a price taker

If, for instance, anyone can decide to start a dairy farm, and if any existing dairy farmer can decide to leave the dairy business, then the dairy industry would satisfy this condition. It should be noted that much of the analysis of competitive firms does not rely on the assumption of free entry and exit because this condition is not necessary for firms to be price takers. But as we will see later in this chapter, entry and exit are often powerful forces shaping the long-run outcome in competitive markets.

THE REVENUE OF A COMPETITIVE FIRM

A firm in a competitive market, like most other firms in the economy, tries to maximize profit, which equals total revenue minus total cost. To see how it does this, we first consider the revenue of a competitive firm. To keep matters concrete, let's consider a specific firm: the Smith Family Dairy Farm.

The Smith Farm produces a quantity of milk Q and sells each unit at the market price P . The farm's total revenue is $P \times Q$. For example, if a gallon of milk sells for \$6 and the farm sells 1,000 gallons, its total revenue is \$6,000.

Because the Smith Farm is small compared to the world market for milk, it takes the price as given by market conditions. This means, in particular, that the price of milk does not depend on the quantity of output that the Smith Farm produces and sells. If the Smiths double the amount of milk they produce, the price of milk remains the same, and their total revenue doubles. As a result, total revenue is proportional to the amount of output.

Table 14-1 shows the revenue for the Smith Family Dairy Farm. The first two columns show the amount of output the farm produces and the price at which it sells its output. The third column is the farm's total revenue. The table assumes that the price of milk is \$6 a gallon, so total revenue is simply \$6 times the number of gallons.

Just as the concepts of *average* and *marginal* were useful in the preceding chapter when analyzing costs, they are also useful when analyzing revenue. To see what these concepts tell us, consider these two questions:

QUANTITY (IN GALLONS)	PRICE	TOTAL REVENUE	AVERAGE REVENUE	MARGINAL REVENUE
(Q)	(P)	($TR = P \times Q$)	($AR = TR/Q$)	($MR = \Delta TR/\Delta Q$)
1	\$6	\$ 6	\$6	\$6
2	6	12	6	6
3	6	18	6	6
4	6	24	6	6
5	6	30	6	6
6	6	36	6	6
7	6	42	6	6
8	6	48	6	6

Table 14-1

TOTAL, AVERAGE, AND
MARGINAL REVENUE FOR A
COMPETITIVE FIRM

- ◆ How much revenue does the farm receive for the typical gallon of milk?
- ◆ How much additional revenue does the farm receive if it increases production of milk by 1 gallon?

average revenue

total revenue divided by the quantity sold

marginal revenue

the change in total revenue from an additional unit sold

The last two columns in Table 14-1 answer these questions.

The fourth column in the table shows **average revenue**, which is total revenue (from the third column) divided by the amount of output (from the first column). Average revenue tells us how much revenue a firm receives for the typical unit sold. In Table 14-1, you can see that average revenue equals \$6, the price of a gallon of milk. This illustrates a general lesson that applies not only to competitive firms but to other firms as well. Total revenue is the price times the quantity ($P \times Q$), and average revenue is total revenue ($P \times Q$) divided by the quantity (Q). Therefore, *for all firms, average revenue equals the price of the good.*

The fifth column shows **marginal revenue**, which is the change in total revenue from the sale of each additional unit of output. In Table 14-1, marginal revenue equals \$6, the price of a gallon of milk. This result illustrates a lesson that applies only to competitive firms. Total revenue is $P \times Q$, and P is fixed for a competitive firm. Therefore, when Q rises by 1 unit, total revenue rises by P dollars. *For competitive firms, marginal revenue equals the price of the good.*

QUICK QUIZ: When a competitive firm doubles the amount it sells, what happens to the price of its output and its total revenue?

PROFIT MAXIMIZATION AND THE COMPETITIVE FIRM'S SUPPLY CURVE

The goal of a competitive firm is to maximize profit, which equals total revenue minus total cost. We have just discussed the firm's revenue, and in the last chapter we discussed the firm's costs. We are now ready to examine how the firm maximizes profit and how that decision leads to its supply curve.

A SIMPLE EXAMPLE OF PROFIT MAXIMIZATION

Let's begin our analysis of the firm's supply decision with the example in Table 14-2. In the first column of the table is the number of gallons of milk the Smith Family Dairy Farm produces. The second column shows the farm's total revenue, which is \$6 times the number of gallons. The third column shows the farm's total cost. Total cost includes fixed costs, which are \$3 in this example, and variable costs, which depend on the quantity produced.

The fourth column shows the farm's profit, which is computed by subtracting total cost from total revenue. If the farm produces nothing, it has a loss of \$3. If it produces 1 gallon, it has a profit of \$1. If it produces 2 gallons, it has a profit of \$4, and so on. To maximize profit, the Smith Farm chooses the quantity that makes profit as large as possible. In this example, profit is maximized when the farm produces 4 or 5 gallons of milk, when the profit is \$7.

QUANTITY (IN GALLONS)	TOTAL REVENUE	TOTAL COST	PROFIT	MARGINAL REVENUE	MARGINAL COST
(Q)	(TR)	(TC)	(TR – TC)	(MR = $\Delta TR/\Delta Q$)	(MC = $\Delta TC/\Delta Q$)
0	\$ 0	\$ 3	–\$3		
1	6	5	1	\$6	\$2
2	12	8	4	6	3
3	18	12	6	6	4
4	24	17	7	6	5
5	30	23	7	6	6
6	36	30	6	6	7
7	42	38	4	6	8
8	48	47	1	6	9

PROFIT MAXIMIZATION: A NUMERICAL EXAMPLE

Table 14-2

There is another way to look at the Smith Farm's decision: The Smiths can find the profit-maximizing quantity by comparing the marginal revenue and marginal cost from each unit produced. The last two columns in Table 14-2 compute marginal revenue and marginal cost from the changes in total revenue and total cost. The first gallon of milk the farm produces has a marginal revenue of \$6 and a marginal cost of \$2; hence, producing that gallon increases profit by \$4 (from –\$3 to \$1). The second gallon produced has a marginal revenue of \$6 and a marginal cost of \$3, so that gallon increases profit by \$3 (from \$1 to \$4). As long as marginal revenue exceeds marginal cost, increasing the quantity produced raises profit. Once the Smith Farm has reached 5 gallons of milk, however, the situation is very different. The sixth gallon would have marginal revenue of \$6 and marginal cost of \$7, so producing it would reduce profit by \$1 (from \$7 to \$6). As a result, the Smiths would not produce beyond 5 gallons.

One of the *Ten Principles of Economics* in Chapter 1 is that rational people think at the margin. We now see how the Smith Family Dairy Farm can apply this principle. If marginal revenue is greater than marginal cost—as it is at 1, 2, or 3 gallons—the Smiths should increase the production of milk. If marginal revenue is less than marginal cost—as it is at 6, 7, or 8 gallons—the Smiths should decrease production. If the Smiths think at the margin and make incremental adjustments to the level of production, they are naturally led to produce the profit-maximizing quantity.

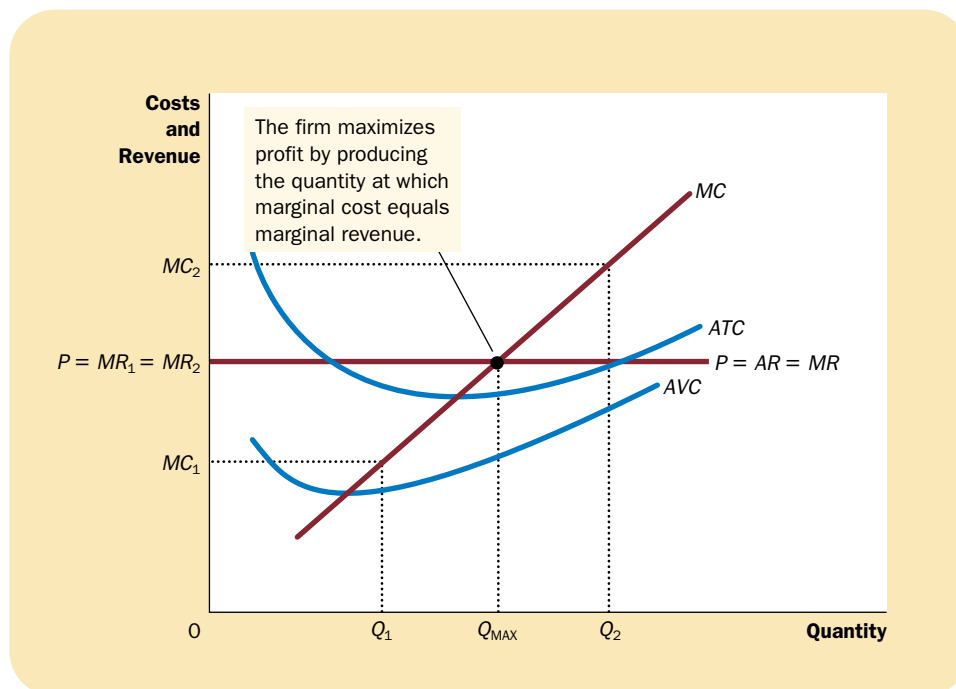


THE MARGINAL-COST CURVE AND THE FIRM'S SUPPLY DECISION

To extend this analysis of profit maximization, consider the cost curves in Figure 14-1. These cost curves have the three features that, as we discussed in Chapter 13, are thought to describe most firms: The marginal-cost curve (MC) is upward sloping. The average-total-cost curve (ATC) is U-shaped. And the marginal-cost curve

Figure 14-1

PROFIT MAXIMIZATION FOR A COMPETITIVE FIRM. This figure shows the marginal-cost curve (MC), the average-total-cost curve (ATC), and the average-variable-cost curve (AVC). It also shows the market price (P), which equals marginal revenue (MR) and average revenue (AR). At the quantity Q_1 , marginal revenue MR_1 exceeds marginal cost MC_1 , so raising production increases profit. At the quantity Q_2 , marginal cost MC_2 is above marginal revenue MR_2 , so reducing production increases profit. The profit-maximizing quantity Q_{MAX} is found where the horizontal price line intersects the marginal-cost curve.



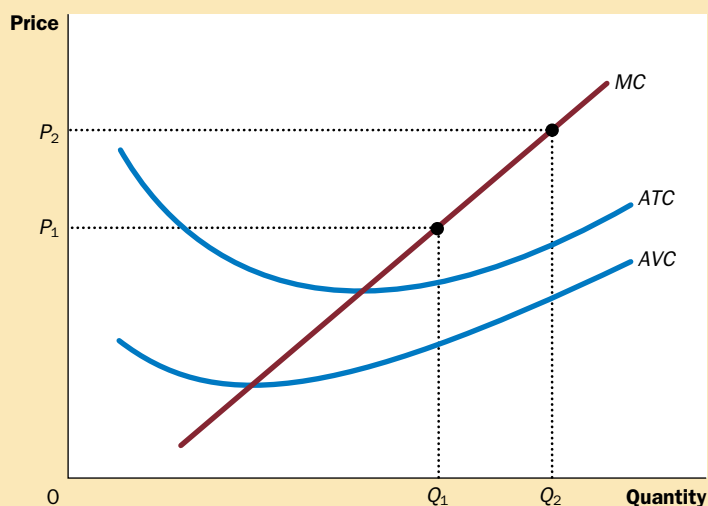
crosses the average-total-cost curve at the minimum of average total cost. The figure also shows a horizontal line at the market price (P). The price line is horizontal because the firm is a price taker: The price of the firm's output is the same regardless of the quantity that the firm decides to produce. Keep in mind that, for a competitive firm, the firm's price equals both its average revenue (AR) and its marginal revenue (MR).

We can use Figure 14-1 to find the quantity of output that maximizes profit. Imagine that the firm is producing at Q_1 . At this level of output, marginal revenue is greater than marginal cost. That is, if the firm raised its level of production and sales by 1 unit, the additional revenue (MR_1) would exceed the additional costs (MC_1). Profit, which equals total revenue minus total cost, would increase. Hence, if marginal revenue is greater than marginal cost, as it is at Q_1 , the firm can increase profit by increasing production.

A similar argument applies when output is at Q_2 . In this case, marginal cost is greater than marginal revenue. If the firm reduced production by 1 unit, the costs saved (MC_2) would exceed the revenue lost (MR_2). Therefore, if marginal revenue is less than marginal cost, as it is at Q_2 , the firm can increase profit by reducing production.

Where do these marginal adjustments to level of production end? Regardless of whether the firm begins with production at a low level (such as Q_1) or at a high level (such as Q_2), the firm will eventually adjust production until the quantity produced reaches Q_{MAX} . This analysis shows a general rule for profit maximization: *At the profit-maximizing level of output, marginal revenue and marginal cost are exactly equal.*

We can now see how the competitive firm decides the quantity of its good to supply to the market. Because a competitive firm is a price taker, its marginal

**Figure 14-2**

MARGINAL COST AS THE COMPETITIVE FIRM'S SUPPLY CURVE. An increase in the price from P_1 to P_2 leads to an increase in the firm's profit-maximizing quantity from Q_1 to Q_2 . Because the marginal-cost curve shows the quantity supplied by the firm at any given price, it is the firm's supply curve.

revenue equals the market price. For any given price, the competitive firm's profit-maximizing quantity of output is found by looking at the intersection of the price with the marginal-cost curve. In Figure 14-1, that quantity of output is Q_{MAX} .

Figure 14-2 shows how a competitive firm responds to an increase in the price. When the price is P_1 , the firm produces quantity Q_1 , which is the quantity that equates marginal cost to the price. When the price rises to P_2 , the firm finds that marginal revenue is now higher than marginal cost at the previous level of output, so the firm increases production. The new profit-maximizing quantity is Q_2 , at which marginal cost equals the new higher price. *In essence, because the firm's marginal-cost curve determines the quantity of the good the firm is willing to supply at any price, it is the competitive firm's supply curve.*

THE FIRM'S SHORT-RUN DECISION TO SHUT DOWN

So far we have been analyzing the question of how much a competitive firm will produce. In some circumstances, however, the firm will decide to shut down and not produce anything at all.

Here we should distinguish between a temporary shutdown of a firm and the permanent exit of a firm from the market. A *shutdown* refers to a short-run decision not to produce anything during a specific period of time because of current market conditions. *Exit* refers to a long-run decision to leave the market. The short-run and long-run decisions differ because most firms cannot avoid their fixed costs in the short run but can do so in the long run. That is, a firm that shuts down temporarily still has to pay its fixed costs, whereas a firm that exits the market saves both its fixed and its variable costs.

For example, consider the production decision that a farmer faces. The cost of the land is one of the farmer's fixed costs. If the farmer decides not to produce any

crops one season, the land lies fallow, and he cannot recover this cost. When making the short-run decision whether to shut down for a season, the fixed cost of land is said to be a *sunk cost*. By contrast, if the farmer decides to leave farming altogether, he can sell the land. When making the long-run decision whether to exit the market, the cost of land is not sunk. (We return to the issue of sunk costs shortly.)

Now let's consider what determines a firm's shutdown decision. If the firm shuts down, it loses all revenue from the sale of its product. At the same time, it saves the variable costs of making its product (but must still pay the fixed costs). Thus, *the firm shuts down if the revenue that it would get from producing is less than its variable costs of production.*

A small bit of mathematics can make this shutdown criterion more useful. If TR stands for total revenue, and VC stands for variable costs, then the firm's decision can be written as

$$\text{Shut down if } TR < VC.$$

The firm shuts down if total revenue is less than variable cost. By dividing both sides of this inequality by the quantity Q , we can write it as

$$\text{Shut down if } TR/Q < VC/Q.$$

Notice that this can be further simplified. TR/Q is total revenue divided by quantity, which is average revenue. As we discussed previously, average revenue for any firm is simply the good's price P . Similarly, VC/Q is average variable cost AVC . Therefore, the firm's shutdown criterion is

$$\text{Shut down if } P < AVC.$$

That is, a firm chooses to shut down if the price of the good is less than the average variable cost of production. This criterion is intuitive: When choosing to produce, the firm compares the price it receives for the typical unit to the average variable cost that it must incur to produce the typical unit. If the price doesn't cover the average variable cost, the firm is better off stopping production altogether. The firm can reopen in the future if conditions change so that price exceeds average variable cost.

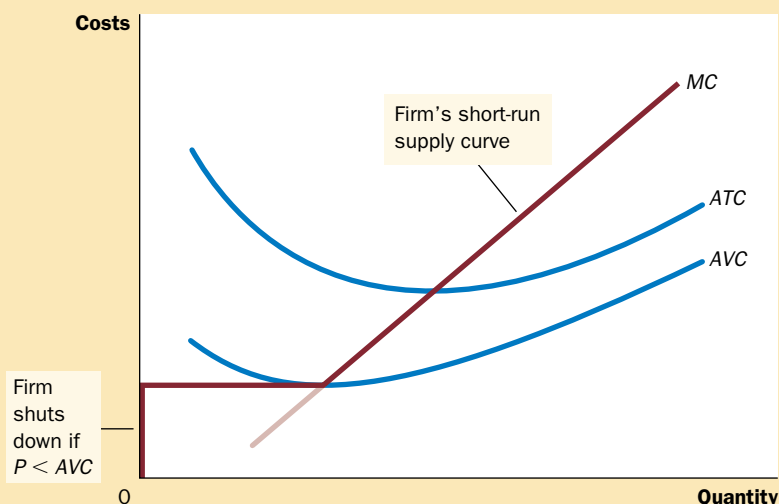
We now have a full description of a competitive firm's profit-maximizing strategy. If the firm produces anything, it produces the quantity at which marginal cost equals the price of the good. Yet if the price is less than average variable cost at that quantity, the firm is better off shutting down and not producing anything. These results are illustrated in Figure 14-3. *The competitive firm's short-run supply curve is the portion of its marginal-cost curve that lies above average variable cost.*

SPILT MILK AND OTHER SUNK COSTS

Sometime in your life, you have probably been told, "Don't cry over spilt milk," or "Let bygones be bygones." These adages hold a deep truth about rational decision-making. Economists say that a cost is a **sunk cost** when it has already been committed and cannot be recovered. In a sense, a sunk cost is the opposite of an opportunity cost: An opportunity cost is what you have to give up if you choose to

sunk cost

a cost that has already been committed and cannot be recovered

Figure 14-3

THE COMPETITIVE FIRM'S SHORT-RUN SUPPLY CURVE. In the short run, the competitive firm's supply curve is its marginal-cost curve (MC) above average variable cost (AVC). If the price falls below average variable cost, the firm is better off shutting down.

do one thing instead of another, whereas a sunk cost cannot be avoided, regardless of the choices you make. Because nothing can be done about sunk costs, you can ignore them when making decisions about various aspects of life, including business strategy.

Our analysis of the firm's shutdown decision is one example of the irrelevance of sunk costs. We assume that the firm cannot recover its fixed costs by temporarily stopping production. As a result, the firm's fixed costs are sunk in the short run, and the firm can safely ignore these costs when deciding how much to produce. The firm's short-run supply curve is the part of the marginal-cost curve that lies above average variable cost, and the size of the fixed cost does not matter for this supply decision.

The irrelevance of sunk costs explains how real businesses make decisions. In the early 1990s, for instance, most of the major airlines reported large losses. In one year, American Airlines, Delta, and USAir each lost more than \$400 million. Yet despite the losses, these airlines continued to sell tickets and fly passengers. At first, this decision might seem surprising: If the airlines were losing money flying planes, why didn't the owners of the airlines just shut down their businesses?

To understand this behavior, we must acknowledge that many of the airlines' costs are sunk in the short run. If an airline has bought a plane and cannot resell it, then the cost of the plane is sunk. The opportunity cost of a flight includes only the variable costs of fuel and the wages of pilots and flight attendants. As long as the total revenue from flying exceeds these variable costs, the airlines should continue operating. And, in fact, they did.

The irrelevance of sunk costs is also important for personal decisions. Imagine, for instance, that you place a \$10 value on seeing a newly released movie. You buy a ticket for \$7, but before entering the theater, you lose the ticket. Should you buy another ticket? Or should you now go home and refuse to pay a total of \$14 to see the movie? The answer is that you should buy another ticket. The benefit of seeing

the movie (\$10) still exceeds the opportunity cost (the \$7 for the second ticket). The \$7 you paid for the lost ticket is a sunk cost. As with spilt milk, there is no point in crying about it.



STAYING OPEN CAN BE PROFITABLE, EVEN WITH MANY TABLES EMPTY.

CASE STUDY NEAR-EMPTY RESTAURANTS AND OFF-SEASON MINIATURE GOLF

Have you ever walked into a restaurant for lunch and found it almost empty? Why, you might have asked, does the restaurant even bother to stay open? It might seem that the revenue from the few customers could not possibly cover the cost of running the restaurant.

In making the decision whether to open for lunch, a restaurant owner must keep in mind the distinction between fixed and variable costs. Many of a restaurant's costs—the rent, kitchen equipment, tables, plates, silverware, and so on—are fixed. Shutting down during lunch would not reduce these costs. In other words, these costs are sunk in the short run. When the owner is deciding whether to serve lunch, only the variable costs—the price of the additional food and the wages of the extra staff—are relevant. The owner shuts down the restaurant at lunchtime only if the revenue from the few lunchtime customers fails to cover the restaurant's variable costs.

An operator of a miniature-golf course in a summer resort community faces a similar decision. Because revenue varies substantially from season to season, the firm must decide when to open and when to close. Once again, the fixed costs—the costs of buying the land and building the course—are irrelevant. The miniature-golf course should be open for business only during those times of year when its revenue exceeds its variable costs.

THE FIRM'S LONG-RUN DECISION TO EXIT OR ENTER A MARKET

The firm's long-run decision to exit the market is similar to its shutdown decision. If the firm exits, it again will lose all revenue from the sale of its product, but now it saves on both fixed and variable costs of production. Thus, *the firm exits the market if the revenue it would get from producing is less than its total costs.*

We can again make this criterion more useful by writing it mathematically. If TR stands for total revenue, and TC stands for total cost, then the firm's criterion can be written as

$$\text{Exit if } TR < TC.$$

The firm exits if total revenue is less than total cost. By dividing both sides of this inequality by quantity Q , we can write it as

$$\text{Exit if } TR/Q < TC/Q.$$

We can simplify this further by noting that TR/Q is average revenue, which equals the price P , and that TC/Q is average total cost ATC . Therefore, the firm's exit criterion is

Exit if $P < ATC$.

That is, a firm chooses to exit if the price of the good is less than the average total cost of production.

A parallel analysis applies to an entrepreneur who is considering starting a firm. The firm will enter the market if such an action would be profitable, which occurs if the price of the good exceeds the average total cost of production. The entry criterion is

Enter if $P > ATC$.

The criterion for entry is exactly the opposite of the criterion for exit.

We can now describe a competitive firm's long-run profit-maximizing strategy. If the firm is in the market, it produces the quantity at which marginal cost equals the price of the good. Yet if the price is less than average total cost at that quantity, the firm chooses to exit (or not enter) the market. These results are illustrated in Figure 14-4. *The competitive firm's long-run supply curve is the portion of its marginal cost curve that lies above average total cost.*

MEASURING PROFIT IN OUR GRAPH FOR THE COMPETITIVE FIRM

As we analyze exit and entry, it is useful to be able to analyze the firm's profit in more detail. Recall that profit equals total revenue (TR) minus total cost (TC):

$$\text{Profit} = TR - TC.$$

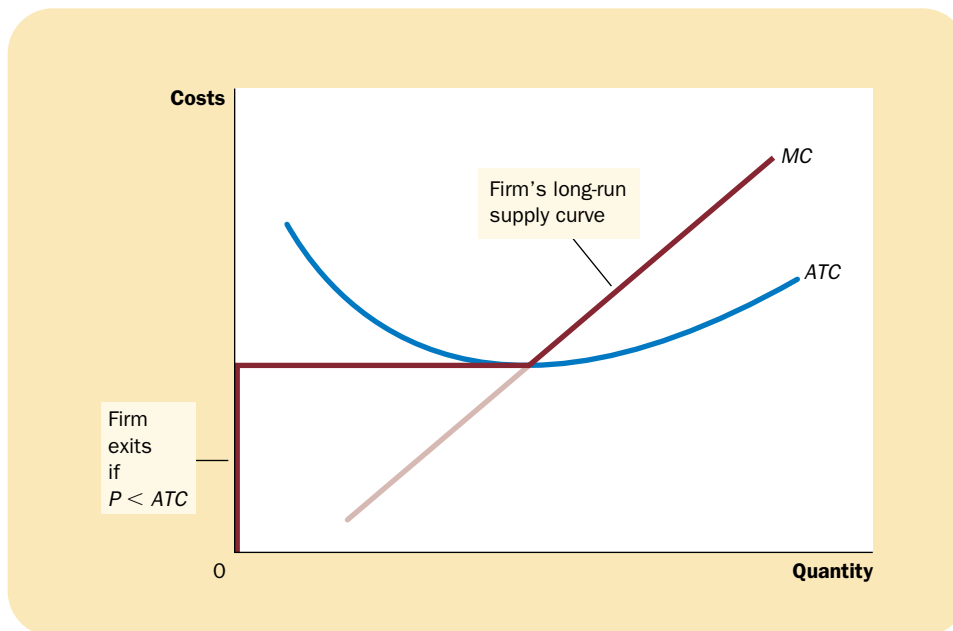


Figure 14-4

THE COMPETITIVE FIRM'S LONG-RUN SUPPLY CURVE. In the long run, the competitive firm's supply curve is its marginal-cost curve (MC) above average total cost (ATC). If the price falls below average total cost, the firm is better off exiting the market.

IN THE NEWS*Entry and Exit in
Transition Economies*

IN THE 1990s, MANY COUNTRIES THAT HAD previously relied on communist theories of central planning tried to make the transition to free-market capitalism. According to this article, Poland succeeded because it encouraged free entry and exit, and Russia failed because it didn't.

**Russia Is Not Poland,
and That's Too Bad**

BY MICHAEL M. WEINSTEIN

Put aside for a moment the frightening crash of the ruble and the collapse of

Russia's stock and bond markets last week. They are symptoms of something larger—a deformed economy in which the Government sets business taxes that few firms ever pay, enterprises promise wages that employees never see, loans go unpaid, people barter with pots, pans and socks, and shady dealing runs rampant.

It didn't have to be this way. The Russians need only look to Poland to behold the better road untraveled. Poland too began the decade saddled with paltry living standards bequeathed by a sclerotic, centrally controlled economy run by discredited Communists. It reached out to the West for help creating monetary, budget, trade and legal regimes, and unlike Russia it followed through with sustained political will. It now ranks among Europe's fastest-growing economies.

Key to Poland's steady success have been two policy decisions, and discussing them helps to illuminate by contrast what is going wrong with Russia.

First, Poland adopted what might be called the Balcerowicz rule, named after Leszek Balcerowicz, the Finance Minister who masterminded Poland's market reforms. Mr. Balcerowicz invited thousands of would-be entrepreneurs to sell, within loose limits, anything they wanted anywhere they wanted at whatever price they wanted. Economists called this liberalization. The Poles called it competition.

The Balcerowicz rule helped break the chokehold of Communist-dominated, state-owned enterprises and Government bureaucracies over economic activity. Also, encouraging small start-ups denies organized crime opportunities for large prey.

When Poland broke away from communism, Western economists had wrung their hands trying to figure out what to do with its sprawling state-owned factories, which operated more like social welfare agencies than production units. The solution, it turned out, was benign neglect. Rather than convert factories, the Poles allowed them to

We can rewrite this definition by multiplying and dividing the right-hand side by Q :

$$\text{Profit} = (TR/Q - TC/Q) \times Q.$$

But note that TR/Q is average revenue, which is the price P , and TC/Q is average total cost ATC . Therefore,

$$\text{Profit} = (P - ATC) \times Q.$$

This way of expressing the firm's profit allows us to measure profit in our graphs.

Panel (a) of Figure 14-5 shows a firm earning positive profit. As we have already discussed, the firm maximizes profit by producing the quantity at which price equals marginal cost. Now look at the shaded rectangle. The height of the rectangle is $P - ATC$, the difference between price and average total cost. The

shriveled. Workers peeled away to set up retail shops and other small enterprises largely free of Government interference.

The second major decision was scarier. Poland forced insolvent firms into bankruptcy, preventing them from draining resources from productive parts of the economy. That also ended a drain on the Federal budget by firms that had to be propped up by one disguised subsidy or another.

There were moments when the post-Communist Government in Russia appeared headed in the same direction. In early 1992, the Yeltsin Government embraced the Balcerowicz rule. Russians were invited to take to the streets and set up kiosks and curbside tables, selling whatever they wanted at whatever price consumers would pay. But then Communist antibodies, in the form of the oligarchs who controlled the state-owned factories and natural resources, were activated. They detected foreign tissue and attacked. Local government buried the Balcerowicz rule, imposing licensing and other requirements and

eventually strangling start-ups. Professor Marshall Goldman of Harvard points to revealing comments by Viktor S. Chernomyrdin, the off-again, on-again Prime Minister whom President Boris N. Yeltsin restored to his post last week. Mr. Chernomyrdin observed that street vendors were an unattractive, chaotic blight on a proud country. The Russian authorities cracked down.

The impact was severe. Anders Aslund, a former adviser to the Russian Government now at the Carnegie Endowment for International Peace, estimates that since the middle of 1994, the number of enterprises in Russia has stagnated. In a typical Western economy, he estimates, there is 1 business for every 10 residents. In Russia, the ratio is 1 for every 55.

By snuffing out start-ups, Russia lost the remarkable device by which Poland drained workers out of worthless factories into units that could produce the goods that people wanted to buy.

Russia not only stifles start-ups; it also props up incompetents. It tolerates

businesses that cannot pay taxes or wages. They survive because of systems of barter and mutual forbearance of loans and taxes. Suppliers engage in round-robin lending by which everyone owes money to someone and no one ever pays up. That too throws a lifeline to insolvent firms.

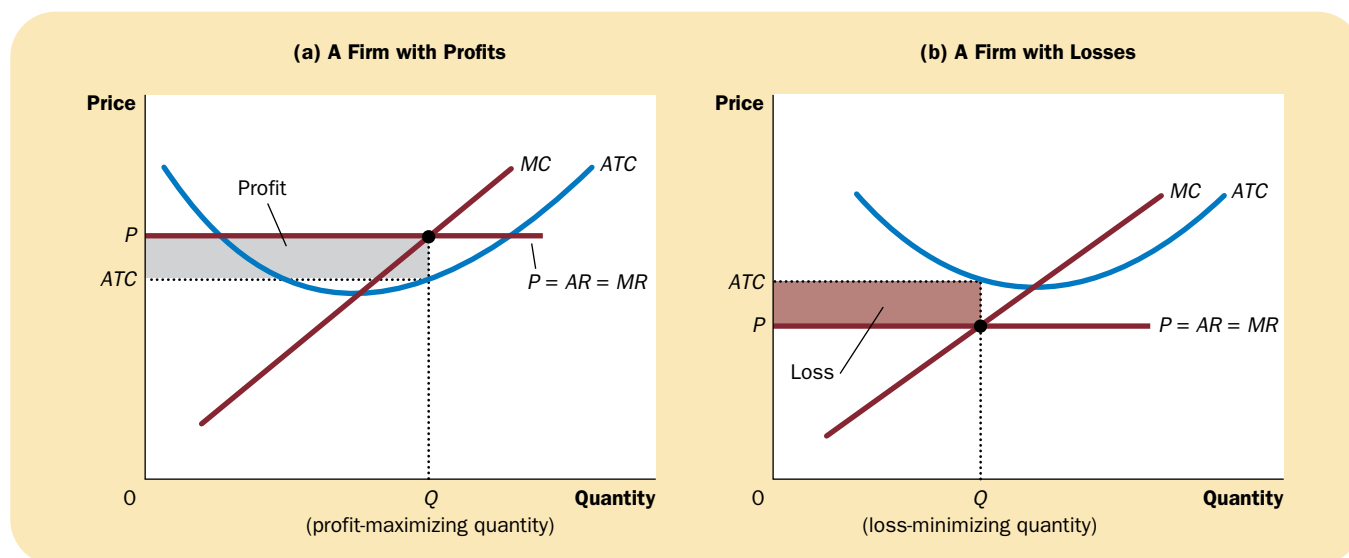
Russian factories continue to churn out steel and other products that no one needs. One measure of the deformity is that Russia is littered with factories employing 10,000 or more workers. In the United States, such factories are a rarity. The effect is to keep alive concerns that chew up \$1.50 worth of resources in order to turn out a product that is worth only \$1 to consumers. Economists call this "negative value added." Ordinary folk call it economic suicide.

SOURCE: *The New York Times*, August 30, 1998, Week in Review, p. 5.

width of the rectangle is Q , the quantity produced. Therefore, the area of the rectangle is $(P - ATC) \times Q$, which is the firm's profit.

Similarly, panel (b) of this figure shows a firm with losses (negative profit). In this case, maximizing profit means minimizing losses, a task accomplished once again by producing the quantity at which price equals marginal cost. Now consider the shaded rectangle. The height of the rectangle is $ATC - P$, and the width is Q . The area is $(ATC - P) \times Q$, which is the firm's loss. Because a firm in this situation is not making enough revenue to cover its average total cost, the firm would choose to exit the market.

QUICK QUIZ: How does the price faced by a profit-maximizing competitive firm compare to its marginal cost? Explain. ♦ When does a profit-maximizing competitive firm decide to shut down?

**Figure 14-5**

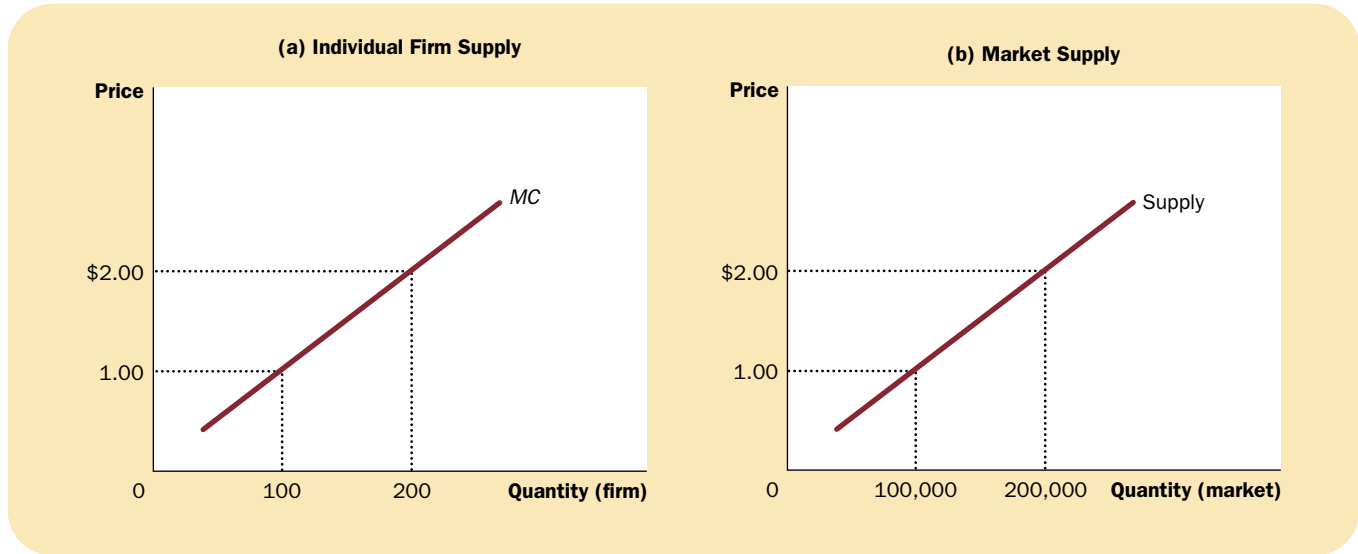
PROFIT AS THE AREA BETWEEN PRICE AND AVERAGE TOTAL COST. The area of the shaded box between price and average total cost represents the firm's profit. The height of this box is price minus average total cost ($P - ATC$), and the width of the box is the quantity of output (Q). In panel (a), price is above average total cost, so the firm has positive profit. In panel (b), price is less than average total cost, so the firm has losses.

THE SUPPLY CURVE IN A COMPETITIVE MARKET

Now that we have examined the supply decision of a single firm, we can discuss the supply curve for a market. There are two cases to consider. First, we examine a market with a fixed number of firms. Second, we examine a market in which the number of firms can change as old firms exit the market and new firms enter. Both cases are important, for each applies over a specific time horizon. Over short periods of time, it is often difficult for firms to enter and exit, so the assumption of a fixed number of firms is appropriate. But over long periods of time, the number of firms can adjust to changing market conditions.

THE SHORT RUN: MARKET SUPPLY WITH A FIXED NUMBER OF FIRMS

Consider first a market with 1,000 identical firms. For any given price, each firm supplies a quantity of output so that its marginal cost equals the price, as shown in panel (a) of Figure 14-6. That is, as long as price is above average variable cost, each firm's marginal-cost curve is its supply curve. The quantity of output supplied to the market equals the sum of the quantities supplied by each of the 1,000 individual firms. Thus, to derive the market supply curve, we add the quantity supplied by each firm in the market. As panel (b) of Figure 14-6 shows, because the



MARKET SUPPLY WITH A FIXED NUMBER OF FIRMS. When the number of firms in the market is fixed, the market supply curve, shown in panel (b), reflects the individual firms' marginal-cost curves, shown in panel (a). Here, in a market of 1,000 firms, the quantity of output supplied to the market is 1,000 times the quantity supplied by each firm.

Figure 14-6

firms are identical, the quantity supplied to the market is 1,000 times the quantity supplied by each firm.

THE LONG RUN: MARKET SUPPLY WITH ENTRY AND EXIT

Now consider what happens if firms are able to enter or exit the market. Let's suppose that everyone has access to the same technology for producing the good and access to the same markets to buy the inputs into production. Therefore, all firms and all potential firms have the same cost curves.

Decisions about entry and exit in a market of this type depend on the incentives facing the owners of existing firms and the entrepreneurs who could start new firms. If firms already in the market are profitable, then new firms will have an incentive to enter the market. This entry will expand the number of firms, increase the quantity of the good supplied, and drive down prices and profits. Conversely, if firms in the market are making losses, then some existing firms will exit the market. Their exit will reduce the number of firms, decrease the quantity of the good supplied, and drive up prices and profits. *At the end of this process of entry and exit, firms that remain in the market must be making zero economic profit.* Recall that we can write a firm's profits as

$$\text{Profit} = (P - ATC) \times Q.$$

This equation shows that an operating firm has zero profit if and only if the price of the good equals the average total cost of producing that good. If price is above average total cost, profit is positive, which encourages new firms to enter. If price

is less than average total cost, profit is negative, which encourages some firms to exit. *The process of entry and exit ends only when price and average total cost are driven to equality.*

This analysis has a surprising implication. We noted earlier in the chapter that competitive firms produce so that price equals marginal cost. We just noted that free entry and exit forces price to equal average total cost. But if price is to equal both marginal cost and average total cost, these two measures of cost must equal each other. Marginal cost and average total cost are equal, however, only when the firm is operating at the minimum of average total cost. Therefore, *the long-run equilibrium of a competitive market with free entry and exit must have firms operating at their efficient scale.*

Panel (a) of Figure 14-7 shows a firm in such a long-run equilibrium. In this figure, price P equals marginal cost MC , so the firm is profit-maximizing. Price also equals average total cost ATC , so profits are zero. New firms have no incentive to enter the market, and existing firms have no incentive to leave the market.

From this analysis of firm behavior, we can determine the long-run supply curve for the market. In a market with free entry and exit, there is only one price consistent with zero profit—the minimum of average total cost. As a result, the long-run market supply curve must be horizontal at this price, as in panel (b) of Figure 14-7. Any price above this level would generate profit, leading to entry and an increase in the total quantity supplied. Any price below this level would generate losses, leading to exit and a decrease in the total quantity supplied. Eventually, the number of firms in the market adjusts so that price equals the minimum of

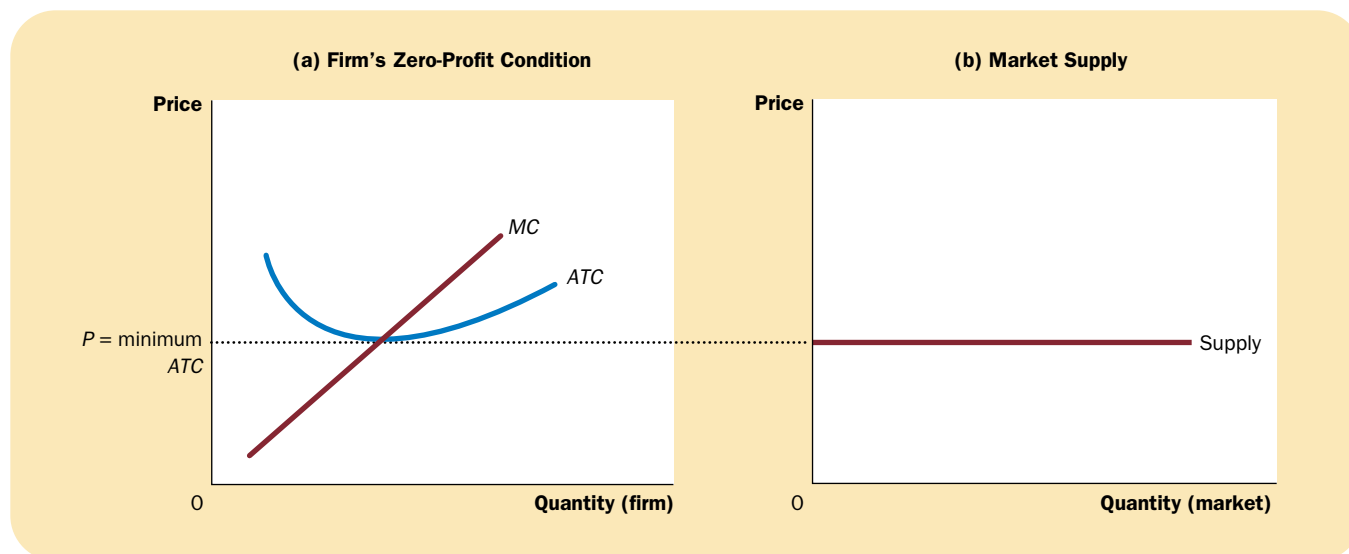


Figure 14-7

MARKET SUPPLY WITH ENTRY AND EXIT. Firms will enter or exit the market until profit is driven to zero. Thus, in the long run, price equals the minimum of average total cost, as shown in panel (a). The number of firms adjusts to ensure that all demand is satisfied at this price. The long-run market supply curve is horizontal at this price, as shown in panel (b).

average total cost, and there are enough firms to satisfy all the demand at this price.

WHY DO COMPETITIVE FIRMS STAY IN BUSINESS IF THEY MAKE ZERO PROFIT?

At first, it might seem odd that competitive firms earn zero profit in the long run. After all, people start businesses to make a profit. If entry eventually drives profit to zero, there might seem to be little reason to stay in business.

To understand the zero-profit condition more fully, recall that profit equals total revenue minus total cost, and that total cost includes all the opportunity costs of the firm. In particular, total cost includes the opportunity cost of the time and money that the firm owners devote to the business. In the zero-profit equilibrium, the firm's revenue must compensate the owners for the time and money that they expend to keep their business going.

Consider an example. Suppose that a farmer had to invest \$1 million to open his farm, which otherwise he could have deposited in a bank to earn \$50,000 a year in interest. In addition, he had to give up another job that would have paid him \$30,000 a year. Then the farmer's opportunity cost of farming includes both the interest he could have earned and the forgone wages—a total of \$80,000. Even if his profit is driven to zero, his revenue from farming compensates him for these opportunity costs.

Keep in mind that accountants and economists measure costs differently. As we discussed in Chapter 13, accountants keep track of explicit costs but usually miss implicit costs. That is, they measure costs that require an outflow of money from the firm, but they fail to include opportunity costs of production that do not involve an outflow of money. As a result, in the zero-profit equilibrium, economic profit is zero, but accounting profit is positive. Our farmer's accountant, for instance, would conclude that the farmer earned an accounting profit of \$80,000, which is enough to keep the farmer in business.



"We're a nonprofit organization—we don't intend to be, but we are!"

A SHIFT IN DEMAND IN THE SHORT RUN AND LONG RUN

Because firms can enter and exit a market in the long run but not in the short run, the response of a market to a change in demand depends on the time horizon. To see this, let's trace the effects of a shift in demand. This analysis will show how a market responds over time, and it will show how entry and exit drive a market to its long-run equilibrium.

Suppose the market for milk begins in long-run equilibrium. Firms are earning zero profit, so price equals the minimum of average total cost. Panel (a) of Figure 14-8 shows the situation. The long-run equilibrium is point A, the quantity sold in the market is Q_1 , and the price is P_1 .

Now suppose scientists discover that milk has miraculous health benefits. As a result, the demand curve for milk shifts outward from D_1 to D_2 , as in panel (b). The short-run equilibrium moves from point A to point B; as a result, the quantity rises from Q_1 to Q_2 , and the price rises from P_1 to P_2 . All of the existing firms respond to the higher price by raising the amount produced. Because each firm's supply curve reflects its marginal-cost curve, how much they each increase production is determined by the marginal-cost curve. In the new, short-run equilibrium, the price of milk exceeds average total cost, so the firms are making positive profit.

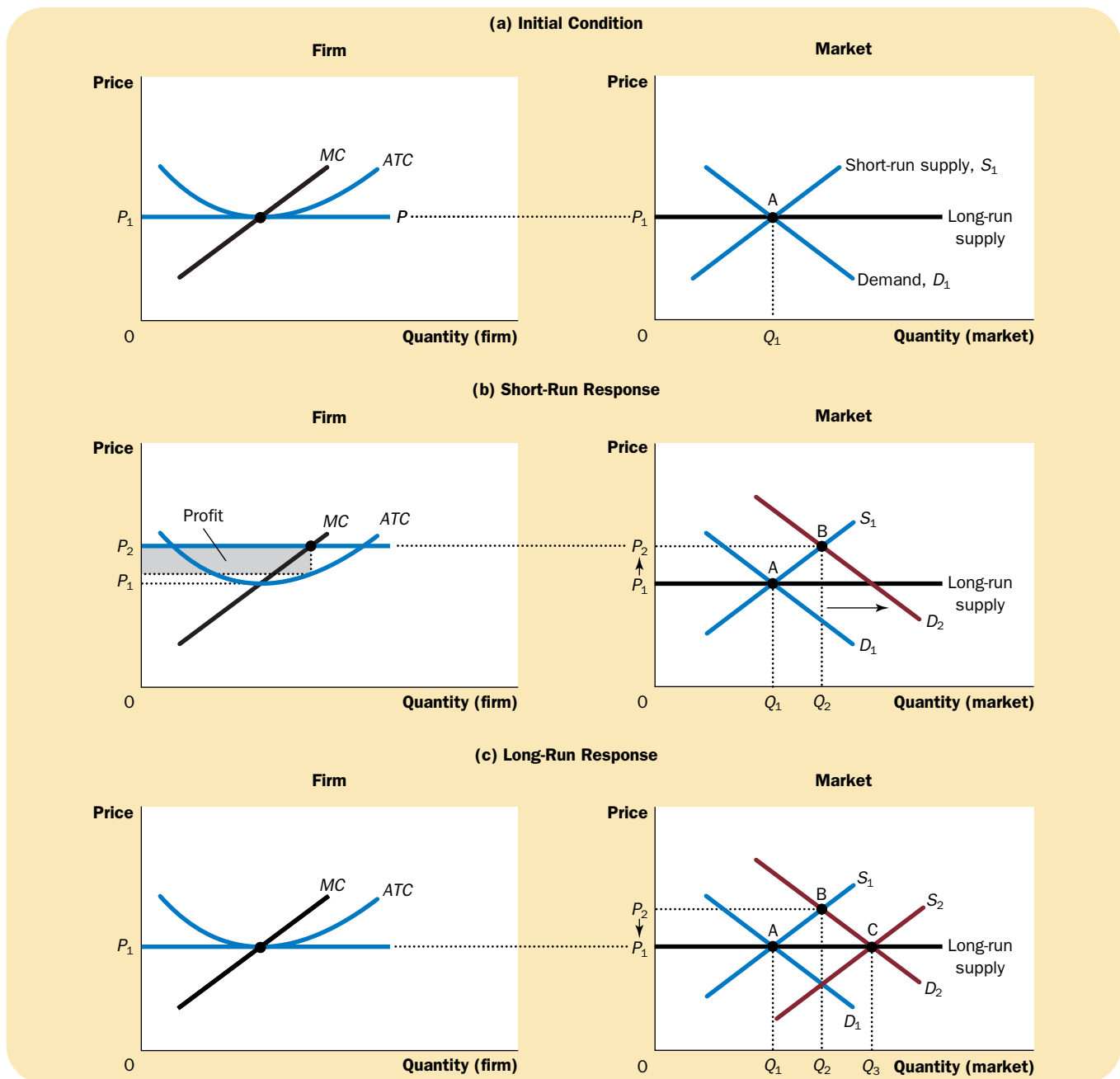
Over time, the profit in this market encourages new firms to enter. Some farmers may switch to milk from other farm products, for example. As the number of firms grows, the short-run supply curve shifts to the right from S_1 to S_2 , as in panel (c), and this shift causes the price of milk to fall. Eventually, the price is driven back down to the minimum of average total cost, profits are zero, and firms stop entering. Thus, the market reaches a new long-run equilibrium, point C. The price of milk has returned to P_1 , but the quantity produced has risen to Q_3 . Each firm is again producing at its efficient scale, but because more firms are in the dairy business, the quantity of milk produced and sold is higher.

WHY THE LONG-RUN SUPPLY CURVE MIGHT SLOPE UPWARD

So far we have seen that entry and exit can cause the long-run market supply curve to be horizontal. The essence of our analysis is that there are a large number of potential entrants, each of which faces the same costs. As a result, the long-run market supply curve is horizontal at the minimum of average total cost. When the demand for the good increases, the long-run result is an increase in the number of firms and in the total quantity supplied, without any change in the price.

There are, however, two reasons that the long-run market supply curve might slope upward. The first is that some resource used in production may be available only in limited quantities. For example, consider the market for farm products. Anyone can choose to buy land and start a farm, but the quantity of land is limited. As more people become farmers, the price of farmland is bid up, which raises the costs of all farmers in the market. Thus, an increase in demand for farm products cannot induce an increase in quantity supplied without also inducing a rise in farmers' costs, which in turn means a rise in price. The result is a long-run market supply curve that is upward sloping, even with free entry into farming.

A second reason for an upward-sloping supply curve is that firms may have different costs. For example, consider the market for painters. Anyone can enter



AN INCREASE IN DEMAND IN THE SHORT RUN AND LONG RUN. The market starts in a long-run equilibrium, shown as point A in panel (a). In this equilibrium, each firm makes zero profit, and the price equals the minimum average total cost. Panel (b) shows what happens in the short run when demand rises from D_1 to D_2 . The equilibrium goes from point A to point B, price rises from P_1 to P_2 , and the quantity sold in the market rises from Q_1 to Q_2 . Because price now exceeds average total cost, firms make profits, which over time encourage new firms to enter the market. This entry shifts the short-run supply curve to the right from S_1 to S_2 , as shown in panel (c). In the new long-run equilibrium, point C, price has returned to P_1 but the quantity sold has increased to Q_3 . Profits are again zero, price is back to the minimum of average total cost, but the market has more firms to satisfy the greater demand.

Figure 14-8

IN THE NEWS*Entry or Overinvestment?*

IN COMPETITIVE MARKETS, STRONG DEMAND leads to high prices and high profits, which then lead to increased entry, falling prices, and falling profits. To economists, these market forces are one reflection of the invisible hand at work. To the business managers, however, new entry and falling profits can seem like a “problem of overinvestment.”

**In Some Industries, Executives
Foresee Tough Times Ahead;
A Key Culprit: High Profits**

BY BERNARD WYSOCKI, JR.
MONTEREY, CALIF.—About 20 executives are huddled in a conference room

with a team of management consultants, and the mood is surprisingly somber.

It's a fine summer day, the stock market is booming, the U.S. economy is in great shape, and some of the companies represented here are posting stronger-than-expected profits. Best of all, perhaps, these lucky executives are just a chip shot away from the famed Pebble Beach golf course. They ought to be euphoric.

Instead, an undertone of concern is evident among these executives from Mobil Corp., Union Carbide Corp. and other capital-intensive companies. In between golf, fine meals and cigars, they hear a sobering message from their hosts.

“I feel like the prophet of doom” is the welcoming line of R. Duane Dickson, a director of Mercer Management Consulting and host of the meeting. “It's our belief that the downturn has started. I can't tell you how far it's going to go. But it could be a very ugly one.”

For two days, the executives and their advisers discuss what they expect in their industries between now and 2000: growing overcapacity, world-wide

product gluts, price wars, shakeouts, and consolidations. . . .

One man who attended the Pebble Beach meeting, Joseph Soviero, a Union Carbide vice president, cites an odd but basic problem in chemicals: the strong profits of the past few years. “The profitability that the industry sees during the good times has always led to overinvesting, and it has this time,” Mr. Soviero says. He adds that the chemicals business cycle is alive and has peaked. At Union Carbide, he says, “we always talk about the cycle” and try to manage it.

So far, demand isn't a big problem. In many industries, it is still growing steadily, though slowly. What is developing is too much supply, stemming from the recurring problem of overinvestment. . . . The next few years will bring fierce competition and falling prices.

SOURCE: *The Wall Street Journal*, August 7, 1997, p. A1.

the market for painting services, but not everyone has the same costs. Costs vary in part because some people work faster than others and in part because some people have better alternative uses of their time than others. For any given price, those with lower costs are more likely to enter than those with higher costs. To increase the quantity of painting services supplied, additional entrants must be encouraged to enter the market. Because these new entrants have higher costs, the price must rise to make entry profitable for them. Thus, the market supply curve for painting services slopes upward even with free entry into the market.

Notice that if firms have different costs, some firms earn profit even in the long run. In this case, the price in the market reflects the average total cost of the *marginal firm*—the firm that would exit the market if the price were any lower. This firm earns zero profit, but firms with lower costs earn positive profit. Entry does not eliminate this profit because would-be entrants have higher costs than firms already in the market. Higher-cost firms will enter only if the price rises, making the market profitable for them.

Thus, for these two reasons, the long-run supply curve in a market may be upward sloping rather than horizontal, indicating that a higher price is necessary to induce a larger quantity supplied. Nonetheless, the basic lesson about entry and exit remains true. *Because firms can enter and exit more easily in the long run than in the short run, the long-run supply curve is typically more elastic than the short-run supply curve.*

QUICK QUIZ: In the long run with free entry and exit, is the price in a market equal to marginal cost, average total cost, both, or neither? Explain with a diagram.

CONCLUSION: BEHIND THE SUPPLY CURVE

We have been discussing the behavior of competitive profit-maximizing firms. You may recall from Chapter 1 that one of the *Ten Principles of Economics* is that rational people think at the margin. This chapter has applied this idea to the competitive firm. Marginal analysis has given us a theory of the supply curve in a competitive market and, as a result, a deeper understanding of market outcomes.



We have learned that when you buy a good from a firm in a competitive market, you can be assured that the price you pay is close to the cost of producing that good. In particular, if firms are competitive and profit-maximizing, the price of a good equals the marginal cost of making that good. In addition, if firms can freely enter and exit the market, the price also equals the lowest possible average total cost of production.

Although we have assumed throughout this chapter that firms are price takers, many of the tools developed here are also useful for studying firms in less competitive markets. In the next three chapters we will examine the behavior of firms with market power. Marginal analysis will again be useful in analyzing these firms, but it will have quite different implications.

Summary

- ◆ Because a competitive firm is a price taker, its revenue is proportional to the amount of output it produces. The price of the good equals both the firm's average revenue and its marginal revenue.
- ◆ To maximize profit, a firm chooses a quantity of output such that marginal revenue equals marginal cost. Because marginal revenue for a competitive firm equals the market price, the firm chooses quantity so that price equals marginal cost. Thus, the firm's marginal cost curve is its supply curve.
- ◆ In the short run when a firm cannot recover its fixed costs, the firm will choose to shut down temporarily if the price of the good is less than average variable cost. In the long run when the firm can recover both fixed and variable costs, it will choose to exit if the price is less than average total cost.
- ◆ In a market with free entry and exit, profits are driven to zero in the long run. In this long-run equilibrium, all firms produce at the efficient scale, price equals the minimum of average total cost, and the number of firms adjusts to satisfy the quantity demanded at this price.
- ◆ Changes in demand have different effects over different time horizons. In the short run, an increase in demand raises prices and leads to profits, and a decrease in

demand lowers prices and leads to losses. But if firms can freely enter and exit the market, then in the long run

the number of firms adjusts to drive the market back to the zero-profit equilibrium.

Key Concepts

competitive market, p. 292
average revenue, p. 294

marginal revenue, p. 294

sunk cost, p. 298

Questions for Review

- What is meant by a competitive firm?
- Draw the cost curves for a typical firm. For a given price, explain how the firm chooses the level of output that maximizes profit.
- Under what conditions will a firm shut down temporarily? Explain.
- Under what conditions will a firm exit a market? Explain.
- Does a firm's price equal marginal cost in the short run, in the long run, or both? Explain.
- Does a firm's price equal the minimum of average total cost in the short run, in the long run, or both? Explain.
- Are market supply curves typically more elastic in the short run or in the long run? Explain.

Problems and Applications

- What are the characteristics of a competitive market? Which of the following drinks do you think is best described by these characteristics? Why aren't the others?
 - tap water
 - bottled water
 - cola
 - beer
- Your roommate's long hours in Chem lab finally paid off—she discovered a secret formula that lets people do an hour's worth of studying in 5 minutes. So far, she's sold 200 doses, and faces the following average-total-cost schedule:

<i>Q</i>	AVERAGE TOTAL COST
199	\$199
200	200
201	201

If a new customer offers to pay your roommate \$300 for one dose, should she make one more? Explain.

- The licorice industry is competitive. Each firm produces 2 million strings of licorice per year. The strings have an average total cost of \$0.20 each, and they sell for \$0.30.
 - What is the marginal cost of a string?

- Is this industry in long-run equilibrium? Why or why not?
- You go out to the best restaurant in town and order a lobster dinner for \$40. After eating half of the lobster, you realize that you are quite full. Your date wants you to finish your dinner, because you can't take it home and because "you've already paid for it." What should you do? Relate your answer to the material in this chapter.
 - Bob's lawn-mowing service is a profit-maximizing, competitive firm. Bob mows lawns for \$27 each. His total cost each day is \$280, of which \$30 is a fixed cost. He mows 10 lawns a day. What can you say about Bob's short-run decision regarding shut down and his long-run decision regarding exit?
 - Consider total cost and total revenue given in the table below:

	QUANTITY							
	0	1	2	3	4	5	6	7
Total cost	\$8	\$9	\$10	\$11	\$13	\$19	\$27	\$37
Total revenue	0	8	16	24	32	40	48	56

- Calculate profit for each quantity. How much should the firm produce to maximize profit?

- b. Calculate marginal revenue and marginal cost for each quantity. Graph them. (Hint: Put the points between whole numbers. For example, the marginal cost between 2 and 3 should be graphed at $2\frac{1}{2}$.) At what quantity do these curves cross? How does this relate to your answer to part (a)?
 - c. Can you tell whether this firm is in a competitive industry? If so, can you tell whether the industry is in a long-run equilibrium?
 7. From *The Wall Street Journal* (July 23, 1991): "Since peaking in 1976, per capita beef consumption in the United States has fallen by 28.6 percent . . . [and] the size of the U.S. cattle herd has shrunk to a 30-year low."
 - a. Using firm and industry diagrams, show the short-run effect of declining demand for beef. Label the diagram carefully and write out in words all of the changes you can identify.
 - b. On a new diagram, show the long-run effect of declining demand for beef. Explain in words.
 8. "High prices traditionally cause expansion in an industry, eventually bringing an end to high prices and manufacturers' prosperity." Explain, using appropriate diagrams.
 9. Suppose the book-printing industry is competitive and begins in a long-run equilibrium.
 - a. Draw a diagram describing the typical firm in the industry.
 - b. Hi-Tech Printing Company invents a new process that sharply reduces the cost of printing books. What happens to Hi-Tech's profits and the price of books in the short run when Hi-Tech's patent prevents other firms from using the new technology?
 - c. What happens in the long run when the patent expires and other firms are free to use the technology?
 10. Many small boats are made of fiberglass, which is derived from crude oil. Suppose that the price of oil rises.
 - a. Using diagrams, show what happens to the cost curves of an individual boat-making firm and to the market supply curve.
 - b. What happens to the profits of boat makers in the short run? What happens to the number of boat makers in the long run?
 11. Suppose that the U.S. textile industry is competitive, and there is no international trade in textiles. In long-run equilibrium, the price per unit of cloth is \$30.
 - a. Describe the equilibrium using graphs for the entire market and for an individual producer.
- Now suppose that textile producers in other countries are willing to sell large quantities of cloth in the United States for only \$25 per unit.
- b. Assuming that U.S. textile producers have large fixed costs, what is the short-run effect of these imports on the quantity produced by an individual producer? What is the short-run effect on profits? Illustrate your answer with a graph.
 - c. What is the long-run effect on the number of U.S. firms in the industry?
 12. Suppose there are 1,000 hot pretzel stands operating in New York City. Each stand has the usual U-shaped average-total-cost curve. The market demand curve for pretzels slopes downward, and the market for pretzels is in long-run competitive equilibrium.
 - a. Draw the current equilibrium, using graphs for the entire market and for an individual pretzel stand.
 - b. Now the city decides to restrict the number of pretzel-stand licenses, reducing the number of stands to only 800. What effect will this action have on the market and on an individual stand that is still operating? Use graphs to illustrate your answer.
 - c. Suppose that the city decides to charge a license fee for the 800 licenses. How will this affect the number of pretzels sold by an individual stand, and the stand's profit? The city wants to raise as much revenue as possible and also wants to ensure that 800 pretzel stands remain in the city. By how much should the city increase the license fee? Show the answer on your graph.
 13. Assume that the gold-mining industry is competitive.
 - a. Illustrate a long-run equilibrium using diagrams for the gold market and for a representative gold mine.
 - b. Suppose that an increase in jewelry demand induces a surge in the demand for gold. Using your diagrams, show what happens in the short run to the gold market and to each existing gold mine.
 - c. If the demand for gold remains high, what would happen to the price over time? Specifically, would the new long-run equilibrium price be above, below, or equal to the short-run equilibrium price in part (b)? Is it possible for the new long-run equilibrium price to be above the original long-run equilibrium price? Explain.
 14. (This problem is challenging.) *The New York Times* (July 1, 1994) reported on a Clinton administration proposal to lift the ban on exporting oil from the North Slope of Alaska. According to the article, the administration said that "the chief effect of the ban has

been to provide California refiners with crude oil cheaper than oil on the world market. . . . The ban created a subsidy for California refiners that had not been passed on to consumers.” Let’s use our analysis of firm behavior to analyze these claims.

- a. Draw the cost curves for a California refiner and for a refiner in another part of the world. Assume that the California refiners have access to inexpensive Alaskan crude oil and that other refiners must buy more expensive crude oil from the Middle East.
- b. All of the refiners produce gasoline for the world gasoline market, which has a single price. In the long-run equilibrium, will this price depend on the costs faced by California producers or the costs faced by other producers? Explain. (Hint: California cannot itself supply the entire world market.) Draw new graphs that illustrate the profits earned by a California refiner and another refiner.
- c. In this model, is there a subsidy to California refiners? Is it passed on to consumers?

15



MONOPOLY

If you own a personal computer, it probably uses some version of Windows, the operating system sold by the Microsoft Corporation. When Microsoft first designed Windows many years ago, it applied for and received a copyright from the government. The copyright gives Microsoft the exclusive right to make and sell copies of the Windows operating system. So if a person wants to buy a copy of Windows, he or she has little choice but to give Microsoft the approximately \$50 that the firm has decided to charge for its product. Microsoft is said to have a *monopoly* in the market for Windows.

Microsoft's business decisions are not well described by the model of firm behavior we developed in Chapter 14. In that chapter, we analyzed competitive markets, in which there are many firms offering essentially identical products, so each firm has little influence over the price it receives. By contrast, a monopoly such as Microsoft has no close competitors and, therefore, can influence the market price of its product. While a competitive firm is a *price taker*, a monopoly firm is a *price maker*.

IN THIS CHAPTER YOU WILL . . .

***Learn why some
markets have only
one seller***

***Analyze how a
monopoly
determines the
quantity to produce
and the price to
charge***

***See how the
monopoly's
decisions affect
economic well-being***

***Consider the
various public
policies aimed at
solving the problem
of monopoly***

***See why monopolies
try to charge
different prices to
different customers***

In this chapter we examine the implications of this market power. We will see that market power alters the relationship between a firm's price and its costs. A competitive firm takes the price of its output as given by the market and then chooses the quantity it will supply so that price equals marginal cost. By contrast, the price charged by a monopoly exceeds marginal cost. This result is clearly true in the case of Microsoft's Windows. The marginal cost of Windows—the extra cost that Microsoft would incur by printing one more copy of the program onto some floppy disks or a CD—is only a few dollars. The market price of Windows is many times marginal cost.

It is perhaps not surprising that monopolies charge high prices for their products. Customers of monopolies might seem to have little choice but to pay whatever the monopoly charges. But, if so, why does a copy of Windows not cost \$500? Or \$5,000? The reason, of course, is that if Microsoft set the price that high, fewer people would buy the product. People would buy fewer computers, switch to other operating systems, or make illegal copies. Monopolies cannot achieve any level of profit they want, because high prices reduce the amount that their customers buy. Although monopolies can control the prices of their goods, their profits are not unlimited.

As we examine the production and pricing decisions of monopolies, we also consider the implications of monopoly for society as a whole. Monopoly firms, like competitive firms, aim to maximize profit. But this goal has very different ramifications for competitive and monopoly firms. As we first saw in Chapter 7, self-interested buyers and sellers in competitive markets are unwittingly led by an invisible hand to promote general economic well-being. By contrast, because monopoly firms are unchecked by competition, the outcome in a market with a monopoly is often not in the best interest of society.



One of the *Ten Principles of Economics* in Chapter 1 is that governments can sometimes improve market outcomes. The analysis in this chapter will shed more light on this principle. As we examine the problems that monopolies raise for society, we will also discuss the various ways in which government policymakers might respond to these problems. The U.S. government, for example, keeps a close eye on Microsoft's business decisions. In 1994, it prevented Microsoft from buying Intuit, a software firm that sells the leading program for personal finance, on the grounds that the combination of Microsoft and Intuit would concentrate too much market power in one firm. Similarly, in 1998, the U.S. Justice Department objected when Microsoft started integrating its Internet browser into its Windows operating system, claiming that this would impede competition from other companies, such as Netscape. This concern led the Justice Department to file suit against Microsoft, the final resolution of which was still unsettled as this book was going to press.

WHY MONOPOLIES ARISE

monopoly

a firm that is the sole seller of a product without close substitutes

A firm is a **monopoly** if it is the sole seller of its product and if its product does not have close substitutes. The fundamental cause of monopoly is *barriers to entry*: A monopoly remains the only seller in its market because other firms cannot enter the market and compete with it. Barriers to entry, in turn, have three main sources:

- ◆ A key resource is owned by a single firm.
- ◆ The government gives a single firm the exclusive right to produce some good or service.
- ◆ The costs of production make a single producer more efficient than a large number of producers.

Let's briefly discuss each of these.

MONOPOLY RESOURCES

The simplest way for a monopoly to arise is for a single firm to own a key resource. For example, consider the market for water in a small town in the Old West. If dozens of town residents have working wells, the competitive model discussed in Chapter 14 describes the behavior of sellers. As a result, the price of a gallon of water is driven to equal the marginal cost of pumping an extra gallon. But if there is only one well in town and it is impossible to get water from anywhere else, then the owner of the well has a monopoly on water. Not surprisingly, the monopolist has much greater market power than any single firm in a competitive market. In the case of a necessity like water, the monopolist could command quite a high price, even if the marginal cost is low.

Although exclusive ownership of a key resource is a potential cause of monopoly, in practice monopolies rarely arise for this reason. Actual economies are large, and resources are owned by many people. Indeed, because many goods are traded internationally, the natural scope of their markets is often worldwide. There are, therefore, few examples of firms that own a resource for which there are no close substitutes.

CASE STUDY THE DEBEERS DIAMOND MONOPOLY

A classic example of a monopoly that arises from the ownership of a key resource is DeBeers, the South African diamond company. DeBeers controls about 80 percent of the world's production of diamonds. Although the firm's share of the market is not 100 percent, it is large enough to exert substantial influence over the market price of diamonds.

How much market power does DeBeers have? The answer depends in part on whether there are close substitutes for its product. If people view emeralds, rubies, and sapphires as good substitutes for diamonds, then DeBeers has relatively little market power. In this case, any attempt by DeBeers to raise the price of diamonds would cause people to switch to other gemstones. But if people view these other stones as very different from diamonds, then DeBeers can exert substantial influence over the price of its product.

DeBeers pays for large amounts of advertising. At first, this decision might seem surprising. If a monopoly is the sole seller of its product, why does it need to advertise? One goal of the DeBeers ads is to differentiate diamonds from other gems in the minds of consumers. When their slogan tells you that "a diamond is forever," you are meant to think that the same is not true of emeralds, rubies, and sapphires. (And notice that the slogan is applied to all diamonds, not just DeBeers diamonds—a sign of DeBeers's monopoly position.) If the ads are



"Rather than a monopoly, we like to consider ourselves 'the only game in town.'"

successful, consumers will view diamonds as unique, rather than as one among many gemstones, and this perception will give DeBeers greater market power.

GOVERNMENT-CREATED MONOPOLIES

In many cases, monopolies arise because the government has given one person or firm the exclusive right to sell some good or service. Sometimes the monopoly arises from the sheer political clout of the would-be monopolist. Kings, for example, once granted exclusive business licenses to their friends and allies. At other times, the government grants a monopoly because doing so is viewed to be in the public interest. For instance, the U.S. government has given a monopoly to a company called Network Solutions, Inc., which maintains the database of all .com, .net, and .org Internet addresses, on the grounds that such data need to be centralized and comprehensive.

The patent and copyright laws are two important examples of how the government creates a monopoly to serve the public interest. When a pharmaceutical company discovers a new drug, it can apply to the government for a patent. If the government deems the drug to be truly original, it approves the patent, which gives the company the exclusive right to manufacture and sell the drug for 20 years. Similarly, when a novelist finishes a book, she can copyright it. The copyright is a government guarantee that no one can print and sell the work without the author's permission. The copyright makes the novelist a monopolist in the sale of her novel.

The effects of patent and copyright laws are easy to see. Because these laws give one producer a monopoly, they lead to higher prices than would occur under competition. But by allowing these monopoly producers to charge higher prices and earn higher profits, the laws also encourage some desirable behavior. Drug companies are allowed to be monopolists in the drugs they discover in order to encourage pharmaceutical research. Authors are allowed to be monopolists in the sale of their books to encourage them to write more and better books.

Thus, the laws governing patents and copyrights have benefits and costs. The benefits of the patent and copyright laws are the increased incentive for creative activity. These benefits are offset, to some extent, by the costs of monopoly pricing, which we examine fully later in this chapter.

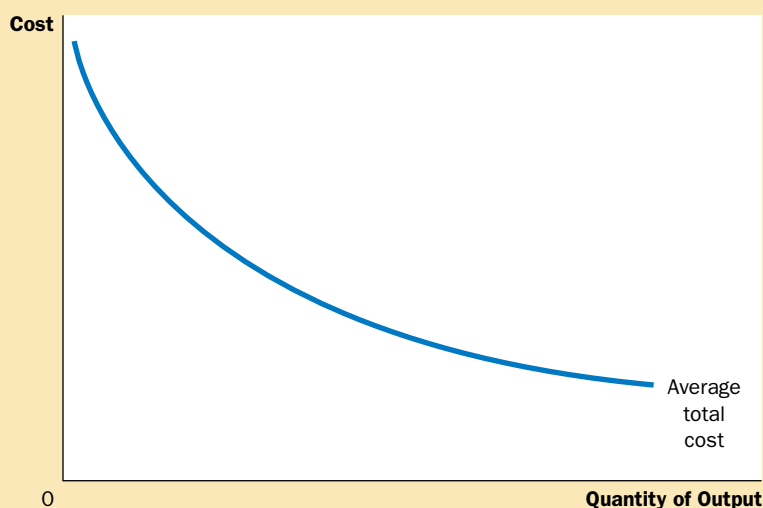
NATURAL MONOPOLIES

natural monopoly

a monopoly that arises because a single firm can supply a good or service to an entire market at a smaller cost than could two or more firms

An industry is a **natural monopoly** when a single firm can supply a good or service to an entire market at a smaller cost than could two or more firms. A natural monopoly arises when there are economies of scale over the relevant range of output. Figure 15-1 shows the average total costs of a firm with economies of scale. In this case, a single firm can produce any amount of output at least cost. That is, for any given amount of output, a larger number of firms leads to less output per firm and higher average total cost.

An example of a natural monopoly is the distribution of water. To provide water to residents of a town, a firm must build a network of pipes throughout the town. If two or more firms were to compete in the provision of this service, each firm would have to pay the fixed cost of building a network. Thus, the average total cost of water is lowest if a single firm serves the entire market.

**Figure 15-1**

ECONOMIES OF SCALE AS A CAUSE OF MONOPOLY. When a firm's average-total-cost curve continually declines, the firm has what is called a natural monopoly. In this case, when production is divided among more firms, each firm produces less, and average total cost rises. As a result, a single firm can produce any given amount at the smallest cost.

We saw other examples of natural monopolies when we discussed public goods and common resources in Chapter 11. We noted in passing that some goods in the economy are excludable but not rival. An example is a bridge used so infrequently that it is never congested. The bridge is excludable because a toll collector can prevent someone from using it. The bridge is not rival because use of the bridge by one person does not diminish the ability of others to use it. Because there is a fixed cost of building the bridge and a negligible marginal cost of additional users, the average total cost of a trip across the bridge (the total cost divided by the number of trips) falls as the number of trips rises. Hence, the bridge is a natural monopoly.

When a firm is a natural monopoly, it is less concerned about new entrants eroding its monopoly power. Normally, a firm has trouble maintaining a monopoly position without ownership of a key resource or protection from the government. The monopolist's profit attracts entrants into the market, and these entrants make the market more competitive. By contrast, entering a market in which another firm has a natural monopoly is unattractive. Would-be entrants know that they cannot achieve the same low costs that the monopolist enjoys because, after entry, each firm would have a smaller piece of the market.

In some cases, the size of the market is one determinant of whether an industry is a natural monopoly. Consider a bridge across a river. When the population is small, the bridge may be a natural monopoly. A single bridge can satisfy the entire demand for trips across the river at lowest cost. Yet as the population grows and the bridge becomes congested, satisfying the entire demand may require two or more bridges across the same river. Thus, as a market expands, a natural monopoly can evolve into a competitive market.

QUICK QUIZ: What are the three reasons that a market might have a monopoly? ♦ Give two examples of monopolies, and explain the reason for each.

HOW MONOPOLIES MAKE PRODUCTION AND PRICING DECISIONS

Now that we know how monopolies arise, we can consider how a monopoly firm decides how much of its product to make and what price to charge for it. The analysis of monopoly behavior in this section is the starting point for evaluating whether monopolies are desirable and what policies the government might pursue in monopoly markets.

MONOPOLY VERSUS COMPETITION

The key difference between a competitive firm and a monopoly is the monopoly's ability to influence the price of its output. A competitive firm is small relative to the market in which it operates and, therefore, takes the price of its output as given by market conditions. By contrast, because a monopoly is the sole producer in its market, it can alter the price of its good by adjusting the quantity it supplies to the market.

One way to view this difference between a competitive firm and a monopoly is to consider the demand curve that each firm faces. When we analyzed profit maximization by competitive firms in Chapter 14, we drew the market price as a horizontal line. Because a competitive firm can sell as much or as little as it wants at this price, the competitive firm faces a horizontal demand curve, as in panel (a) of Figure 15-2. In effect, because the competitive firm sells a product with many

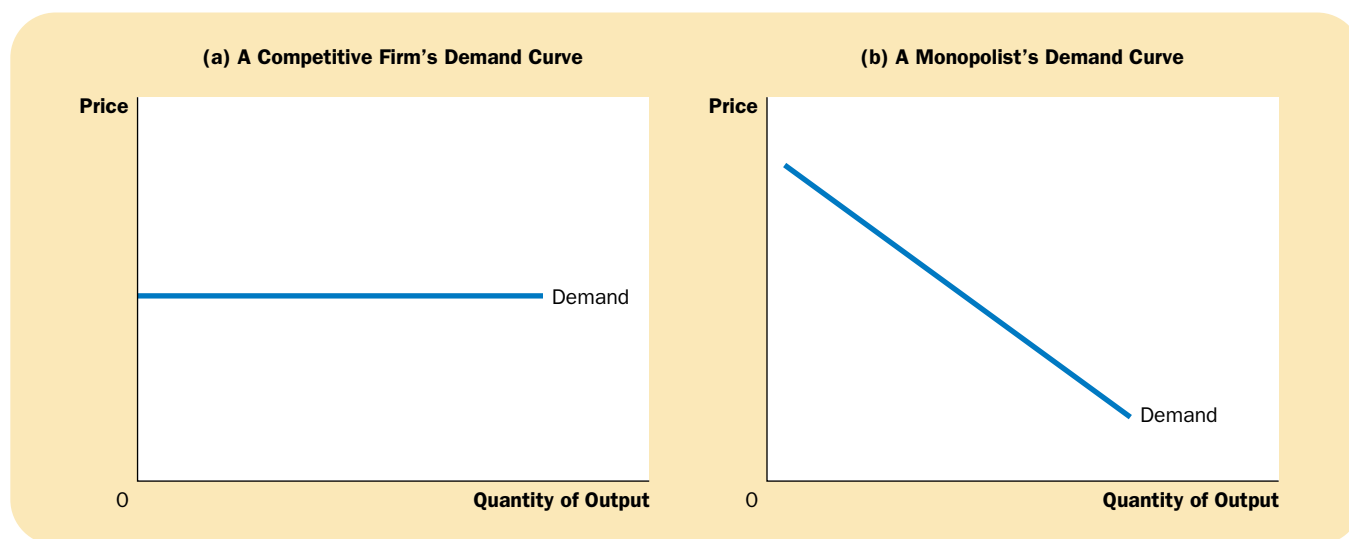


Figure 15-2

DEMAND CURVES FOR COMPETITIVE AND MONOPOLY FIRMS. Because competitive firms are price takers, they in effect face horizontal demand curves, as in panel (a). Because a monopoly firm is the sole producer in its market, it faces the downward-sloping market demand curve, as in panel (b). As a result, the monopoly has to accept a lower price if it wants to sell more output.

perfect substitutes (the products of all the other firms in its market), the demand curve that any one firm faces is perfectly elastic.

By contrast, because a monopoly is the sole producer in its market, its demand curve is the market demand curve. Thus, the monopolist's demand curve slopes downward for all the usual reasons, as in panel (b) of Figure 15-2. If the monopolist raises the price of its good, consumers buy less of it. Looked at another way, if the monopolist reduces the quantity of output it sells, the price of its output increases.

The market demand curve provides a constraint on a monopoly's ability to profit from its market power. A monopolist would prefer, if it were possible, to charge a high price and sell a large quantity at that high price. The market demand curve makes that outcome impossible. In particular, the market demand curve describes the combinations of price and quantity that are available to a monopoly firm. By adjusting the quantity produced (or, equivalently, the price charged), the monopolist can choose any point on the demand curve, but it cannot choose a point off the demand curve.

What point on the demand curve will the monopolist choose? As with competitive firms, we assume that the monopolist's goal is to maximize profit. Because the firm's profit is total revenue minus total costs, our next task in explaining monopoly behavior is to examine a monopolist's revenue.

A MONOPOLY'S REVENUE

Consider a town with a single producer of water. Table 15-1 shows how the monopoly's revenue might depend on the amount of water produced.

The first two columns show the monopolist's demand schedule. If the monopolist produces 1 gallon of water, it can sell that gallon for \$10. If it produces

QUANTITY OF WATER	PRICE	TOTAL REVENUE	AVERAGE REVENUE	MARGINAL REVENUE
(Q)	(P)	($TR = P \times Q$)	($AR = TR/Q$)	($MR = \Delta TR/\Delta Q$)
0 gallons	\$11	\$ 0	—	\$10
1	10	10	\$10	8
2	9	18	9	6
3	8	24	8	4
4	7	28	7	2
5	6	30	6	0
6	5	30	5	-2
7	4	28	4	-4
8	3	24	3	

A MONOPOLY'S TOTAL, AVERAGE, AND MARGINAL REVENUE

Table 15-1

2 gallons, it must lower the price to \$9 in order to sell both gallons. And if it produces 3 gallons, it must lower the price to \$8. And so on. If you graphed these two columns of numbers, you would get a typical downward-sloping demand curve.

The third column of the table presents the monopolist's *total revenue*. It equals the quantity sold (from the first column) times the price (from the second column). The fourth column computes the firm's *average revenue*, the amount of revenue the firm receives per unit sold. We compute average revenue by taking the number for total revenue in the third column and dividing it by the quantity of output in the first column. As we discussed in Chapter 14, average revenue always equals the price of the good. This is true for monopolists as well as for competitive firms.

The last column of Table 15-1 computes the firm's *marginal revenue*, the amount of revenue that the firm receives for each additional unit of output. We compute marginal revenue by taking the change in total revenue when output increases by 1 unit. For example, when the firm is producing 3 gallons of water, it receives total revenue of \$24. Raising production to 4 gallons increases total revenue to \$28. Thus, marginal revenue is \$28 minus \$24, or \$4.

Table 15-1 shows a result that is important for understanding monopoly behavior: *A monopolist's marginal revenue is always less than the price of its good.* For example, if the firm raises production of water from 3 to 4 gallons, it will increase total revenue by only \$4, even though it will be able to sell each gallon for \$7. For a monopoly, marginal revenue is lower than price because a monopoly faces a downward-sloping demand curve. To increase the amount sold, a monopoly firm must lower the price of its good. Hence, to sell the fourth gallon of water, the monopolist must get less revenue for each of the first three gallons.

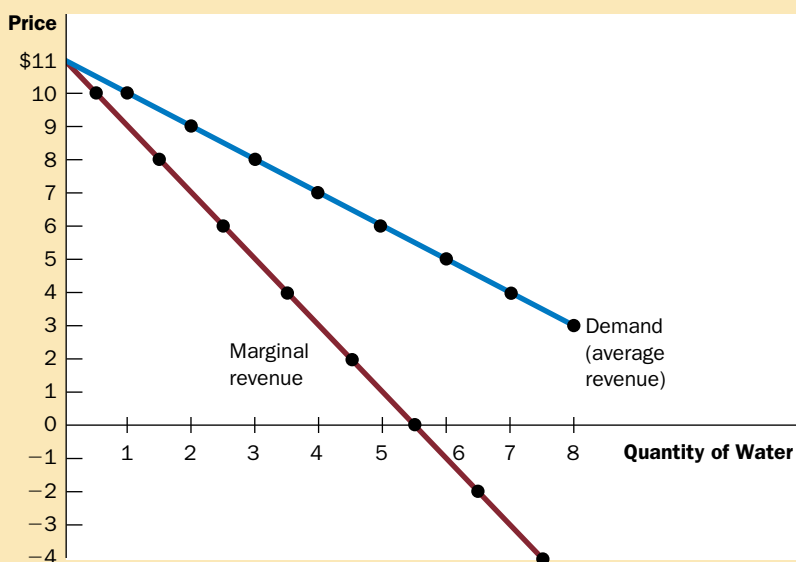
Marginal revenue is very different for monopolies from what it is for competitive firms. When a monopoly increases the amount it sells, it has two effects on total revenue ($P \times Q$):

- ◆ *The output effect:* More output is sold, so Q is higher.
- ◆ *The price effect:* The price falls, so P is lower.

Because a competitive firm can sell all it wants at the market price, there is no price effect. When it increases production by 1 unit, it receives the market price for that unit, and it does not receive any less for the amount it was already selling. That is, because the competitive firm is a price taker, its marginal revenue equals the price of its good. By contrast, when a monopoly increases production by 1 unit, it must reduce the price it charges for every unit it sells, and this cut in price reduces revenue on the units it was already selling. As a result, a monopoly's marginal revenue is less than its price.

Figure 15-3 graphs the demand curve and the marginal-revenue curve for a monopoly firm. (Because the firm's price equals its average revenue, the demand curve is also the average-revenue curve.) These two curves always start at the same point on the vertical axis because the marginal revenue of the first unit sold equals the price of the good. But, for the reason we just discussed, the monopolist's marginal revenue is less than the price of the good. Thus, a monopoly's marginal-revenue curve lies below its demand curve.

You can see in the figure (as well as in Table 15-1) that marginal revenue can even become negative. Marginal revenue is negative when the price effect on

**Figure 15-3**

DEMAND AND MARGINAL-REVENUE CURVES FOR A MONOPOLY. The demand curve shows how the quantity affects the price of the good. The marginal-revenue curve shows how the firm's revenue changes when the quantity increases by 1 unit. Because the price on *all* units sold must fall if the monopoly increases production, marginal revenue is always less than the price.

revenue is greater than the output effect. In this case, when the firm produces an extra unit of output, the price falls by enough to cause the firm's total revenue to decline, even though the firm is selling more units.

PROFIT MAXIMIZATION

Now that we have considered the revenue of a monopoly firm, we are ready to examine how such a firm maximizes profit. Recall from Chapter 1 that one of the *Ten Principles of Economics* is that rational people think at the margin. This lesson is as true for monopolists as it is for competitive firms. Here we apply the logic of marginal analysis to the monopolist's problem of deciding how much to produce.

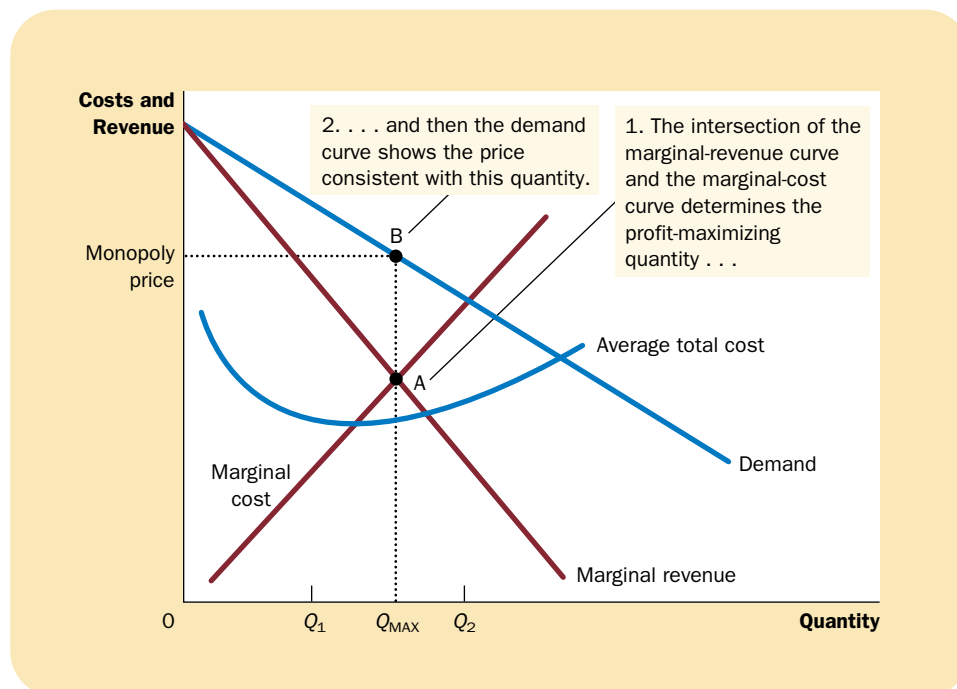
Figure 15-4 graphs the demand curve, the marginal-revenue curve, and the cost curves for a monopoly firm. All these curves should seem familiar: The demand and marginal-revenue curves are like those in Figure 15-3, and the cost curves are like those we introduced in Chapter 13 and used to analyze competitive firms in Chapter 14. These curves contain all the information we need to determine the level of output that a profit-maximizing monopolist will choose.

Suppose, first, that the firm is producing at a low level of output, such as Q_1 . In this case, marginal cost is less than marginal revenue. If the firm increased production by 1 unit, the additional revenue would exceed the additional costs, and profit would rise. Thus, when marginal cost is less than marginal revenue, the firm can increase profit by producing more units.



Figure 15-4

PROFIT MAXIMIZATION FOR A MONOPOLY. A monopoly maximizes profit by choosing the quantity at which marginal revenue equals marginal cost (point A). It then uses the price that will induce consumers to buy that quantity (point B).



A similar argument applies at high levels of output, such as Q_2 . In this case, marginal cost is greater than marginal revenue. If the firm reduced production by 1 unit, the costs saved would exceed the revenue lost. Thus, if marginal cost is greater than marginal revenue, the firm can raise profit by reducing production.

In the end, the firm adjusts its level of production until the quantity reaches Q_{MAX} , at which marginal revenue equals marginal cost. Thus, the monopolist's profit-maximizing quantity of output is determined by the intersection of the marginal-revenue curve and the marginal-cost curve. In Figure 15-4, this intersection occurs at point A.

You might recall from Chapter 14 that competitive firms also choose the quantity of output at which marginal revenue equals marginal cost. In following this rule for profit maximization, competitive firms and monopolies are alike. But there is also an important difference between these types of firm: The marginal revenue of a competitive firm equals its price, whereas the marginal revenue of a monopoly is less than its price. That is,

$$\begin{array}{ll} \text{For a competitive firm:} & P = MR = MC. \\ \text{For a monopoly firm:} & P > MR = MC. \end{array}$$

The equality of marginal revenue and marginal cost at the profit-maximizing quantity is the same for both types of firm. What differs is the relationship of the price to marginal revenue and marginal cost.

How does the monopoly find the profit-maximizing price for its product? The demand curve answers this question, for the demand curve relates the amount that customers are willing to pay to the quantity sold. Thus, after the monopoly firm chooses the quantity of output that equates marginal revenue and marginal

FYI**Why a Monopoly Does Not Have a Supply Curve**

You may have noticed that we have analyzed the price in a monopoly market using the market demand curve and the firm's cost curves. We have not made any mention of the market supply curve. By contrast, when we analyzed prices in competitive markets beginning in Chapter 4, the two most important words were always *supply* and *demand*.

What happened to the supply curve? Although monopoly firms make decisions about what quantity to supply (in the way described in this chapter), a monopoly does not have a supply curve. A supply curve

tells us the quantity that firms choose to supply at any given price. This concept makes sense when we are analyzing competitive firms, which are price takers. But a monopoly firm is a price maker, not a price taker. It is not meaningful to ask what such a firm would produce at any price because the firm sets the price at the same time it chooses the quantity to supply.

Indeed, the monopolist's decision about how much to supply is impossible to separate from the demand curve it faces. The shape of the demand curve determines the shape of the marginal-revenue curve, which in turn determines the monopolist's profit-maximizing quantity. In a competitive market, supply decisions can be analyzed without knowing the demand curve, but that is not true in a monopoly market. Therefore, we never talk about a monopoly's supply curve.

cost, it uses the demand curve to find the price consistent with that quantity. In Figure 15-4, the profit-maximizing price is found at point B.

We can now see a key difference between markets with competitive firms and markets with a monopoly firm: *In competitive markets, price equals marginal cost. In monopolized markets, price exceeds marginal cost.* As we will see in a moment, this finding is crucial to understanding the social cost of monopoly.

A MONOPOLY'S PROFIT

How much profit does the monopoly make? To see the monopoly's profit, recall that profit equals total revenue (TR) minus total costs (TC):

$$\text{Profit} = TR - TC.$$

We can rewrite this as

$$\text{Profit} = (TR/Q - TC/Q) \times Q.$$

TR/Q is average revenue, which equals the price P , and TC/Q is average total cost ATC . Therefore,

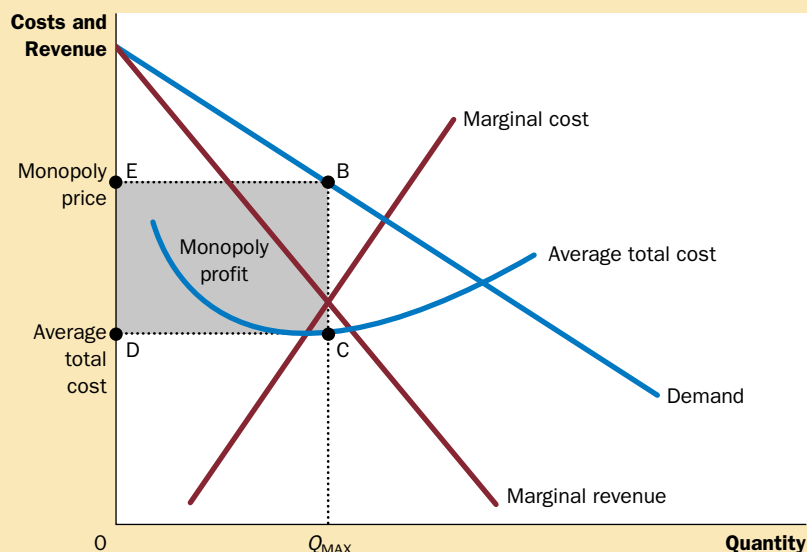
$$\text{Profit} = (P - ATC) \times Q.$$

This equation for profit (which is the same as the profit equation for competitive firms) allows us to measure the monopolist's profit in our graph.

Consider the shaded box in Figure 15-5. The height of the box (the segment BC) is price minus average total cost, $P - ATC$, which is the profit on the typical unit sold. The width of the box (the segment DC) is the quantity sold Q_{MAX} . Therefore, the area of this box is the monopoly firm's total profit.

Figure 15-5

THE MONOPOLIST'S PROFIT.
 The area of the box BCDE equals the profit of the monopoly firm. The height of the box (BC) is price minus average total cost, which equals profit per unit sold. The width of the box (DC) is the number of units sold.



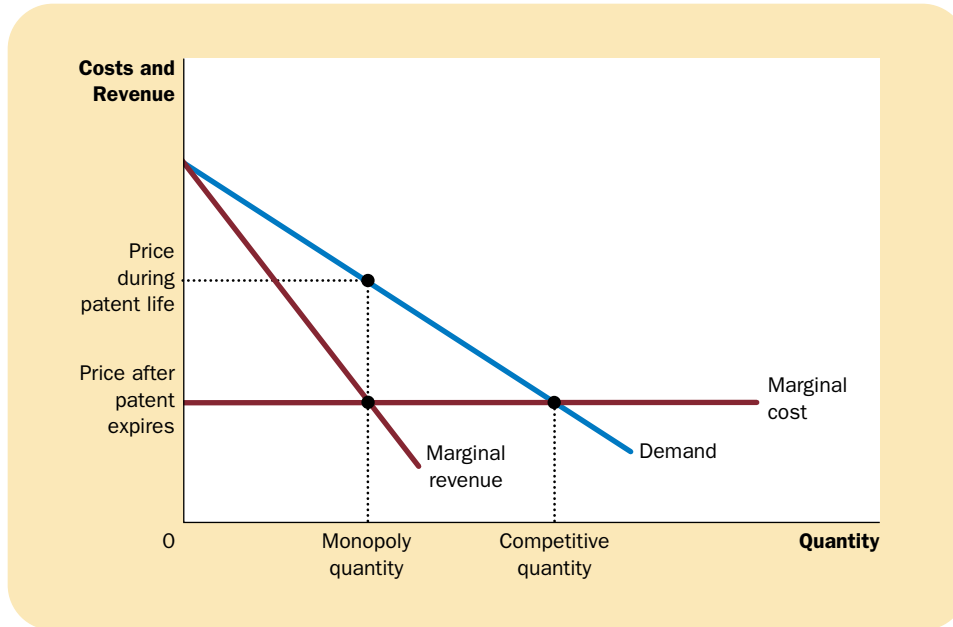
CASE STUDY MONOPOLY DRUGS VERSUS GENERIC DRUGS

According to our analysis, prices are determined quite differently in monopolized markets from the way they are in competitive markets. A natural place to test this theory is the market for pharmaceutical drugs because this market takes on both market structures. When a firm discovers a new drug, patent laws give the firm a monopoly on the sale of that drug. But eventually the firm's patent runs out, and any company can make and sell the drug. At that time, the market switches from being monopolistic to being competitive.

What should happen to the price of a drug when the patent runs out? Figure 15-6 shows the market for a typical drug. In this figure, the marginal cost of producing the drug is constant. (This is approximately true for many drugs.) During the life of the patent, the monopoly firm maximizes profit by producing the quantity at which marginal revenue equals marginal cost and charging a price well above marginal cost. But when the patent runs out, the profit from making the drug should encourage new firms to enter the market. As the market becomes more competitive, the price should fall to equal marginal cost.

Experience is, in fact, consistent with our theory. When the patent on a drug expires, other companies quickly enter and begin selling so-called generic products that are chemically identical to the former monopolist's brand-name product. And just as our analysis predicts, the price of the competitively produced generic drug is well below the price that the monopolist was charging.

The expiration of a patent, however, does not cause the monopolist to lose all its market power. Some consumers remain loyal to the brand-name drug, perhaps out of fear that the new generic drugs are not actually the same as the drug they have been using for years. As a result, the former monopolist can continue to charge a price at least somewhat above the price charged by its new competitors.

**Figure 15-6**

THE MARKET FOR DRUGS. When a patent gives a firm a monopoly over the sale of a drug, the firm charges the monopoly price, which is well above the marginal cost of making the drug. When the patent on a drug runs out, new firms enter the market, making it more competitive. As a result, the price falls from the monopoly price to marginal cost.

QUICK QUIZ: Explain how a monopolist chooses the quantity of output to produce and the price to charge.

THE WELFARE COST OF MONOPOLY

Is monopoly a good way to organize a market? We have seen that a monopoly, in contrast to a competitive firm, charges a price above marginal cost. From the standpoint of consumers, this high price makes monopoly undesirable. At the same time, however, the monopoly is earning profit from charging this high price. From the standpoint of the owners of the firm, the high price makes monopoly very desirable. Is it possible that the benefits to the firm's owners exceed the costs imposed on consumers, making monopoly desirable from the standpoint of society as a whole?

We can answer this question using the type of analysis we first saw in Chapter 7. As in that chapter, we use total surplus as our measure of economic well-being. Recall that total surplus is the sum of consumer surplus and producer surplus. Consumer surplus is consumers' willingness to pay for a good minus the amount they actually pay for it. Producer surplus is the amount producers receive for a good minus their costs of producing it. In this case, there is a single producer: the monopolist.

You might already be able to guess the result of this analysis. In Chapter 7 we concluded that the equilibrium of supply and demand in a competitive market is not only a natural outcome but a desirable one. In particular, the invisible hand of the market leads to an allocation of resources that makes total surplus as large as it can be. Because a monopoly leads to an allocation of resources different from that in a competitive market, the outcome must, in some way, fail to maximize total economic well-being.

THE DEADWEIGHT LOSS

We begin by considering what the monopoly firm would do if it were run by a benevolent social planner. The social planner cares not only about the profit earned by the firm's owners but also about the benefits received by the firm's consumers. The planner tries to maximize total surplus, which equals producer surplus (profit) plus consumer surplus. Keep in mind that total surplus equals the value of the good to consumers minus the costs of making the good incurred by the monopoly producer.

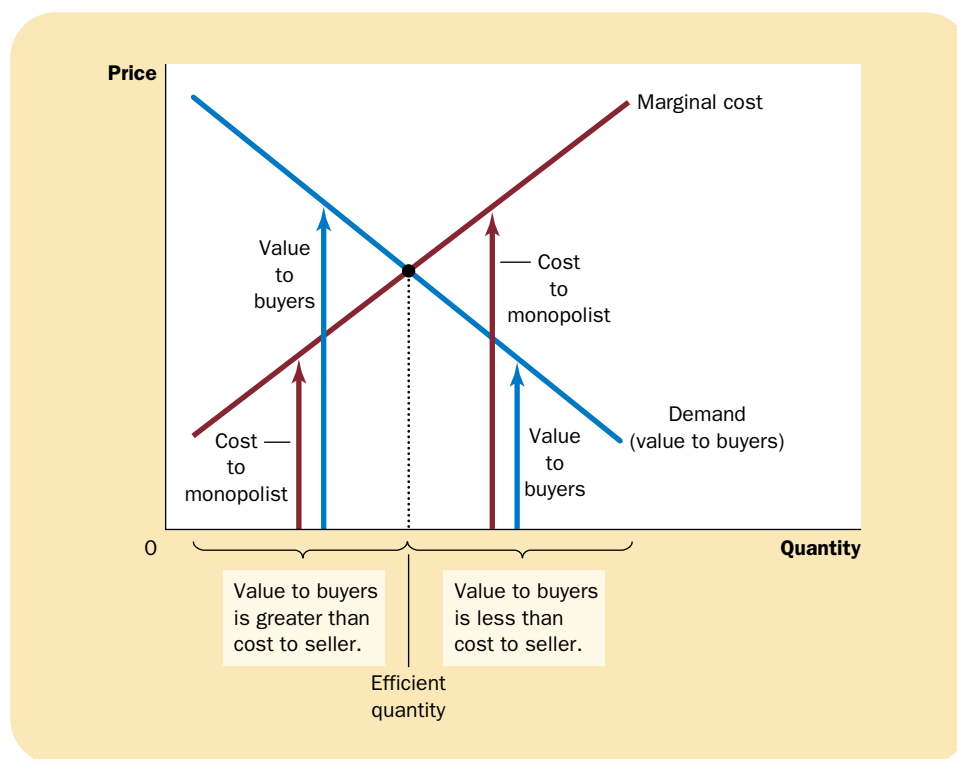
Figure 15-7 analyzes what level of output a benevolent social planner would choose. The demand curve reflects the value of the good to consumers, as measured by their willingness to pay for it. The marginal-cost curve reflects the costs of the monopolist. Thus, the socially efficient quantity is found where the demand curve and the marginal-cost curve intersect. Below this quantity, the value to consumers exceeds the marginal cost of providing the good, so increasing output would raise total surplus. Above this quantity, the marginal cost exceeds the value to consumers, so decreasing output would raise total surplus.

If the social planner were running the monopoly, the firm could achieve this efficient outcome by charging the price found at the intersection of the demand and marginal-cost curves. Thus, like a competitive firm and unlike a profit-maximizing monopoly, a social planner would charge a price equal to marginal cost. Because this price would give consumers an accurate signal about the cost of producing the good, consumers would buy the efficient quantity.

We can evaluate the welfare effects of monopoly by comparing the level of output that the monopolist chooses to the level of output that a social planner

Figure 15-7

THE EFFICIENT LEVEL OF OUTPUT. A benevolent social planner who wanted to maximize total surplus in the market would choose the level of output where the demand curve and marginal-cost curve intersect. Below this level, the value of the good to the marginal buyer (as reflected in the demand curve) exceeds the marginal cost of making the good. Above this level, the value to the marginal buyer is less than marginal cost.



would choose. As we have seen, the monopolist chooses to produce and sell the quantity of output at which the marginal-revenue and marginal-cost curves intersect; the social planner would choose the quantity at which the demand and marginal-cost curves intersect. Figure 15-8 shows the comparison. *The monopolist produces less than the socially efficient quantity of output.*

We can also view the inefficiency of monopoly in terms of the monopolist's price. Because the market demand curve describes a negative relationship between the price and quantity of the good, a quantity that is inefficiently low is equivalent to a price that is inefficiently high. When a monopolist charges a price above marginal cost, some potential consumers value the good at more than its marginal cost but less than the monopolist's price. These consumers do not end up buying the good. Because the value these consumers place on the good is greater than the cost of providing it to them, this result is inefficient. Thus, monopoly pricing prevents some mutually beneficial trades from taking place.

Just as we measured the inefficiency of taxes with the deadweight-loss triangle in Chapter 8, we can similarly measure the inefficiency of monopoly. Figure 15-8 shows the deadweight loss. Recall that the demand curve reflects the value to consumers and the marginal-cost curve reflects the costs to the monopoly producer. Thus, the area of the deadweight-loss triangle between the demand curve and the marginal-cost curve equals the total surplus lost because of monopoly pricing.

The deadweight loss caused by monopoly is similar to the deadweight loss caused by a tax. Indeed, a monopolist is like a private tax collector. As we saw in Chapter 8, a tax on a good places a wedge between consumers' willingness to pay (as reflected in the demand curve) and producers' costs (as reflected in the supply curve). Because a monopoly exerts its market power by charging a price above marginal cost, it places a similar wedge. In both cases, the wedge causes the quantity sold to fall short of the social optimum. The difference between the two cases is that the government gets the revenue from a tax, whereas a private firm gets the monopoly profit.

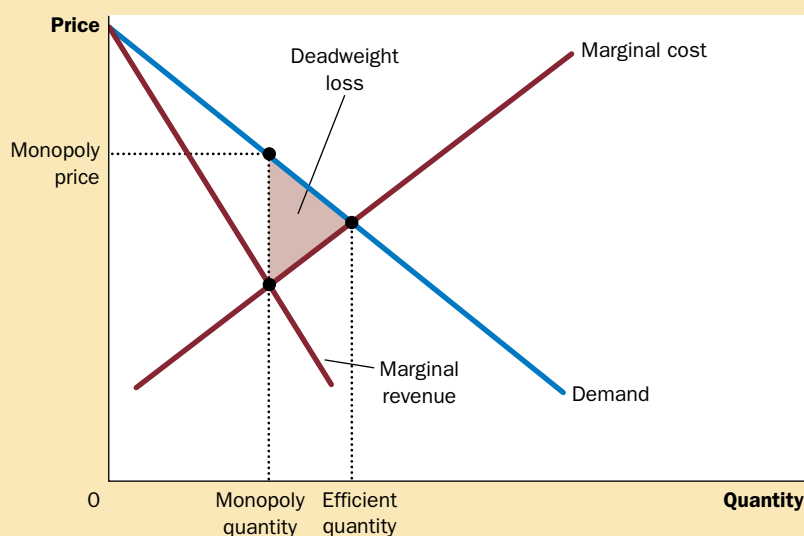


Figure 15-8

THE INEFFICIENCY OF MONOPOLY. Because a monopoly charges a price above marginal cost, not all consumers who value the good at more than its cost buy it. Thus, the quantity produced and sold by a monopoly is below the socially efficient level. The deadweight loss is represented by the area of the triangle between the demand curve (which reflects the value of the good to consumers) and the marginal-cost curve (which reflects the costs of the monopoly producer).

THE MONOPOLY'S PROFIT: A SOCIAL COST?

It is tempting to decry monopolies for “profiteering” at the expense of the public. And, indeed, a monopoly firm does earn a higher profit by virtue of its market power. According to the economic analysis of monopoly, however, the firm’s profit is not in itself necessarily a problem for society.

Welfare in a monopolized market, like all markets, includes the welfare of both consumers and producers. Whenever a consumer pays an extra dollar to a producer because of a monopoly price, the consumer is worse off by a dollar, and the producer is better off by the same amount. This transfer from the consumers of the good to the owners of the monopoly does not affect the market’s total surplus—the sum of consumer and producer surplus. In other words, the monopoly profit itself does not represent a shrinkage in the size of the economic pie; it merely represents a bigger slice for producers and a smaller slice for consumers. Unless consumers are for some reason more deserving than producers—a judgment that goes beyond the realm of economic efficiency—the monopoly profit is not a social problem.

The problem in a monopolized market arises because the firm produces and sells a quantity of output below the level that maximizes total surplus. The deadweight loss measures how much the economic pie shrinks as a result. This inefficiency is connected to the monopoly’s high price: Consumers buy fewer units when the firm raises its price above marginal cost. But keep in mind that the profit earned on the units that continue to be sold is not the problem. The problem stems from the inefficiently low quantity of output. Put differently, if the high monopoly price did not discourage some consumers from buying the good, it would raise producer surplus by exactly the amount it reduced consumer surplus, leaving total surplus the same as could be achieved by a benevolent social planner.

There is, however, a possible exception to this conclusion. Suppose that a monopoly firm has to incur additional costs to maintain its monopoly position. For example, a firm with a government-created monopoly might need to hire lobbyists to convince lawmakers to continue its monopoly. In this case, the monopoly may use up some of its monopoly profits paying for these additional costs. If so, the social loss from monopoly includes both these costs and the deadweight loss resulting from a price above marginal cost.

QUICK QUIZ: How does a monopolist’s quantity of output compare to the quantity of output that maximizes total surplus?

PUBLIC POLICY TOWARD MONOPOLIES

We have seen that monopolies, in contrast to competitive markets, fail to allocate resources efficiently. Monopolies produce less than the socially desirable quantity of output and, as a result, charge prices above marginal cost. Policymakers in the government can respond to the problem of monopoly in one of four ways:

- ◆ By trying to make monopolized industries more competitive
- ◆ By regulating the behavior of the monopolies

- ◆ By turning some private monopolies into public enterprises
- ◆ By doing nothing at all

INCREASING COMPETITION WITH ANTITRUST LAWS

If Coca-Cola and Pepsico wanted to merge, the deal would be closely examined by the federal government before it went into effect. The lawyers and economists in the Department of Justice might well decide that a merger between these two large soft drink companies would make the U.S. soft drink market substantially less competitive and, as a result, would reduce the economic well-being of the country as a whole. If so, the Justice Department would challenge the merger in court, and if the judge agreed, the two companies would not be allowed to merge. It is precisely this kind of challenge that prevented software giant Microsoft from buying Intuit in 1994.

The government derives this power over private industry from the antitrust laws, a collection of statutes aimed at curbing monopoly power. The first and most important of these laws was the Sherman Antitrust Act, which Congress passed in 1890 to reduce the market power of the large and powerful “trusts” that were viewed as dominating the economy at the time. The Clayton Act, passed in 1914, strengthened the government’s powers and authorized private lawsuits. As the U.S. Supreme Court once put it, the antitrust laws are “a comprehensive charter of economic liberty aimed at preserving free and unfettered competition as the rule of trade.”



“But if we do merge with Amalgamated, we’ll have enough resources to fight the anti-trust violation caused by the merger.”

The antitrust laws give the government various ways to promote competition. They allow the government to prevent mergers, such as our hypothetical merger between Coca-Cola and Pepsico. They also allow the government to break up companies. For example, in 1984 the government split up AT&T, the large telecommunications company, into eight smaller companies. Finally, the antitrust laws prevent companies from coordinating their activities in ways that make markets less competitive; we will discuss some of these uses of the antitrust laws in Chapter 16.

Antitrust laws have costs as well as benefits. Sometimes companies merge not to reduce competition but to lower costs through more efficient joint production. These benefits from mergers are sometimes called *synergies*. For example, many U.S. banks have merged in recent years and, by combining operations, have been able to reduce administrative staff. If antitrust laws are to raise social welfare, the government must be able to determine which mergers are desirable and which are not. That is, it must be able to measure and compare the social benefit from synergies to the social costs of reduced competition. Critics of the antitrust laws are skeptical that the government can perform the necessary cost-benefit analysis with sufficient accuracy.

REGULATION

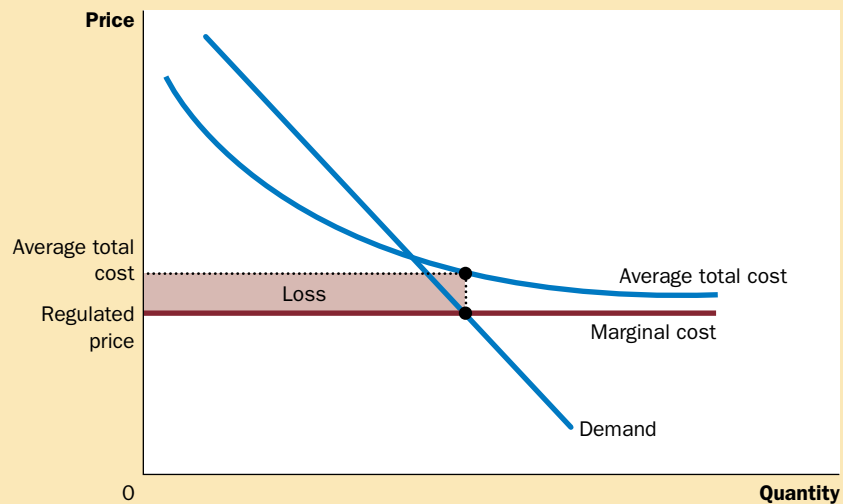
Another way in which the government deals with the problem of monopoly is by regulating the behavior of monopolists. This solution is common in the case of natural monopolies, such as water and electric companies. These companies are not allowed to charge any price they want. Instead, government agencies regulate their prices.

What price should the government set for a natural monopoly? This question is not as easy as it might at first appear. One might conclude that the price should equal the monopolist's marginal cost. If price equals marginal cost, customers will buy the quantity of the monopolist's output that maximizes total surplus, and the allocation of resources will be efficient.

There are, however, two practical problems with marginal-cost pricing as a regulatory system. The first is illustrated in Figure 15-9. Natural monopolies, by definition, have declining average total cost. As we discussed in Chapter 13, when average total cost is declining, marginal cost is less than average total cost. If regulators are to set price equal to marginal cost, that price will be less than the firm's average total cost, and the firm will lose money. Instead of charging such a low price, the monopoly firm would just exit the industry.

Regulators can respond to this problem in various ways, none of which is perfect. One way is to subsidize the monopolist. In essence, the government picks up the losses inherent in marginal-cost pricing. Yet to pay for the subsidy, the government needs to raise money through taxation, which involves its own deadweight losses. Alternatively, the regulators can allow the monopolist to charge a price higher than marginal cost. If the regulated price equals average total cost, the monopolist earns exactly zero economic profit. Yet average-cost pricing leads to deadweight losses, because the monopolist's price no longer reflects the marginal cost of producing the good. In essence, average-cost pricing is like a tax on the good the monopolist is selling.

The second problem with marginal-cost pricing as a regulatory system (and with average-cost pricing as well) is that it gives the monopolist no incentive to

Figure 15-9

MARGINAL-COST PRICING FOR A NATURAL MONOPOLY. Because a natural monopoly has declining average total cost, marginal cost is less than average total cost. Therefore, if regulators require a natural monopoly to charge a price equal to marginal cost, price will be below average total cost, and the monopoly will lose money.

reduce costs. Each firm in a competitive market tries to reduce its costs because lower costs mean higher profits. But if a regulated monopolist knows that regulators will reduce prices whenever costs fall, the monopolist will not benefit from lower costs. In practice, regulators deal with this problem by allowing monopolists to keep some of the benefits from lower costs in the form of higher profit, a practice that requires some departure from marginal-cost pricing.

PUBLIC OWNERSHIP

The third policy used by the government to deal with monopoly is public ownership. That is, rather than regulating a natural monopoly that is run by a private firm, the government can run the monopoly itself. This solution is common in many European countries, where the government owns and operates utilities such as the telephone, water, and electric companies. In the United States, the government runs the Postal Service. The delivery of ordinary First Class mail is often thought to be a natural monopoly.

Economists usually prefer private to public ownership of natural monopolies. The key issue is how the ownership of the firm affects the costs of production. Private owners have an incentive to minimize costs as long as they reap part of the benefit in the form of higher profit. If the firm's managers are doing a bad job of keeping costs down, the firm's owners will fire them. By contrast, if the government bureaucrats who run a monopoly do a bad job, the losers are the customers and taxpayers, whose only recourse is the political system. The bureaucrats may become a special-interest group and attempt to block cost-reducing reforms. Put simply, as a way of ensuring that firms are well run, the voting booth is less reliable than the profit motive.

IN THE NEWS*Public Transport and
Private Enterprise*

IN MANY CITIES, THE MASS TRANSIT SYSTEM of buses and subways is a monopoly run by the local government. But is this the best system?

Man with a Van

BY JOHN TIERNEY

Vincent Cummins looks out from his van with the wary eyes of a hardened criminal. It is quiet this evening in downtown Brooklyn . . . too quiet. "Watch my back for me!" he barks into the microphone of his C.B. radio, addressing a fellow outlaw in a van who just drove by him on Livingston Street. He looks left and right. No police cars in sight. None of the usual unmarked cars, either. Cummins pauses for a second—he has heard on the C.B. that cops have just busted two other drivers—but he can't stop himself. "Watch my back!" he repeats into the

radio as he ruthlessly pulls over to the curb.

Five seconds later, evil triumphs. A middle-aged woman with a shopping bag climbs into the van . . . and Cummins drives off with impunity! His new victim and the other passengers laugh when asked why they're riding this illegal jitney. What fool would pay \$1.50 to stand on the bus or subway when you're guaranteed a seat here for \$1? Unlike bus drivers, the van drivers make change and accept bills, and the vans run more frequently at every hour of the day. "It takes me an hour to get home if I use the bus," explains Cynthia Peters, a nurse born in Trinidad. "When I'm working late, it's very scary waiting in the dark for the bus and then walking the three blocks home. With Vincent's van, I get home in less than half an hour. He takes me right to the door and waits until I get inside."

Cummins would prefer not to be an outlaw. A native of Barbados, he has been driving his van full time ever since an injury forced him to give up his job as a machinist. "I could be collecting disability," he says, "but it's better to work." He met Federal requirements to run an interstate van service, then spent years trying to get approval to operate in the city. His application, which included more than 900 supporting statements



VINCENT CUMMINS: OUTLAW ENTREPRENEUR

from riders, business groups, and church leaders, was approved by the City Taxi and Limousine Commission as well as by the Department of Transportation. Mayor Giuliani supported him. But this summer the City Council rejected his application for a license, as it has rejected most applications over the past four years, which is why thousands of illegal drivers in Brooklyn and Queens are dodging the police.

DOING NOTHING

Each of the foregoing policies aimed at reducing the problem of monopoly has drawbacks. As a result, some economists argue that it is often best for the government not to try to remedy the inefficiencies of monopoly pricing. Here is the assessment of economist George Stigler, who won the Nobel Prize for his work in industrial organization, writing in the *Fortune Encyclopedia of Economics*:

A famous theorem in economics states that a competitive enterprise economy will produce the largest possible income from a given stock of resources. No real economy meets the exact conditions of the theorem, and all real economies will

Council members claim they're trying to prevent vans from causing accidents and traffic problems, although no one who rides the vans takes these protestations seriously. Vans with accredited and insured drivers like Cummins are no more dangerous or disruptive than taxis. The only danger they pose is to the public transit monopoly, whose union leaders have successfully led the campaign against them.

The van drivers have refuted two modern urban myths: that mass transit must lose money and that it must be a public enterprise. Entrepreneurs like Cummins are thriving today in other cities—Seoul and Buenos Aires rely entirely on private, profitable bus companies—and they once made New York the world leader in mass transit. The first horsecars and elevated trains were developed here by private companies. The first subway was partly financed with a loan from the city, but it was otherwise a private operation, built and run quite profitably with the fare set at a nickel—the equivalent of less than a dollar today.

Eventually though, New York's politicians drove most private transit companies out of business by refusing to adjust the fare for inflation. When the enterprises lost money in the 1920's, Mayor John Hylan offered to teach them efficient

management. If the city ran the subway, he promised, it would make money while preserving the nickel fare and freeing New Yorkers from "serfdom" and "dictatorship" of the "grasping transportation monopolies." But expenses soared as soon as government merged the private systems into a true monopoly. The fare, which remained a nickel through seven decades of private transit, has risen 2,900 percent under public management—and today the Metropolitan Transportation Authority still manages to lose about \$2 per ride. Meanwhile, a jitney driver can provide better service at lower prices and still make a profit.

"Transit could be profitable again if entrepreneurs are given a chance," says Daniel B. Klein, an economist at Santa Clara University in California and the co-author of *Curb Rights*, a new book from the Brookings Institution on mass transit. "Government has demonstrated that it has no more business producing transit than producing cornflakes. It should concentrate instead on establishing new rules to foster competition." To encourage private operators to make a long-term investment in regular service along a route, the Brookings researchers recommend selling them exclusive "curb rights" to pick up passengers waiting at certain stops along the route. That way

part-time opportunists couldn't swoop in to steal regular customers from a long-term operator. But to encourage competition, at other corners along the route there should also be common stops where passengers could be picked up by any licensed jitney or bus.

Elements of this system already exist where jitneys have informally established their own stops separate from the regular buses, but the City Council is trying to eliminate these competitors. Besides denying licenses to new drivers like Cummins, the Council has forbidden veteran drivers with licenses to operate on bus routes. Unless these restrictions are overturned in court—a suit on the drivers' behalf has been filed by the Institute for Justice, a public-interest law firm in Washington—the vans can compete only by breaking the law. At this very moment, despite the best efforts of the police and the Transport Workers Union, somewhere in New York a serial predator like Cummins is luring another unsuspecting victim. He may even be making change for a \$5 bill.

SOURCE: *The New York Times Magazine*, August 10, 1997, p. 22.

fall short of the ideal economy—a difference called "market failure." In my view, however, the degree of "market failure" for the American economy is much smaller than the "political failure" arising from the imperfections of economic policies found in real political systems.

As this quotation makes clear, determining the proper role of the government in the economy requires judgments about politics as well as economics.

QUICK QUIZ: Describe the ways policymakers can respond to the inefficiencies caused by monopolies. List a potential problem with each of these policy responses.

PRICE DISCRIMINATION

price discrimination

the business practice of selling the same good at different prices to different customers

So far we have been assuming that the monopoly firm charges the same price to all customers. Yet in many cases firms try to sell the same good to different customers for different prices, even though the costs of producing for the two customers are the same. This practice is called **price discrimination**.

Before discussing the behavior of a price-discriminating monopolist, we should note that price discrimination is not possible when a good is sold in a competitive market. In a competitive market, there are many firms selling the same good at the market price. No firm is willing to charge a lower price to any customer because the firm can sell all it wants at the market price. And if any firm tried to charge a higher price to a customer, that customer would buy from another firm. For a firm to price discriminate, it must have some market power.

A PARABLE ABOUT PRICING

To understand why a monopolist would want to price discriminate, let's consider a simple example. Imagine that you are the president of Readalot Publishing Company. Readalot's best-selling author has just written her latest novel. To keep things simple, let's imagine that you pay the author a flat \$2 million for the exclusive rights to publish the book. Let's also assume that the cost of printing the book is zero. Readalot's profit, therefore, is the revenue it gets from selling the book minus the \$2 million it has paid to the author. Given these assumptions, how would you, as Readalot's president, decide what price to charge for the book?

Your first step in setting the price is to estimate what the demand for the book is likely to be. Readalot's marketing department tells you that the book will attract two types of readers. The book will appeal to the author's 100,000 die-hard fans. These fans will be willing to pay as much as \$30 for the book. In addition, the book will appeal to about 400,000 less enthusiastic readers who will be willing to pay up to \$5 for the book.

What price maximizes Readalot's profit? There are two natural prices to consider: \$30 is the highest price Readalot can charge and still get the 100,000 die-hard fans, and \$5 is the highest price it can charge and still get the entire market of 500,000 potential readers. It is a matter of simple arithmetic to solve Readalot's problem. At a price of \$30, Readalot sells 100,000 copies, has revenue of \$3 million, and makes profit of \$1 million. At a price of \$5, it sells 500,000 copies, has revenue of \$2.5 million, and makes profit of \$500,000. Thus, Readalot maximizes profit by charging \$30 and forgoing the opportunity to sell to the 400,000 less enthusiastic readers.

Notice that Readalot's decision causes a deadweight loss. There are 400,000 readers willing to pay \$5 for the book, and the marginal cost of providing it to them is zero. Thus, \$2 million of total surplus is lost when Readalot charges the higher price. This deadweight loss is the usual inefficiency that arises whenever a monopolist charges a price above marginal cost.

Now suppose that Readalot's marketing department makes an important discovery: These two groups of readers are in separate markets. All the die-hard fans live in Australia, and all the other readers live in the United States. Moreover, it is

difficult for readers in one country to buy books in the other. How does this discovery affect Readalot's marketing strategy?

In this case, the company can make even more profit. To the 100,000 Australian readers, it can charge \$30 for the book. To the 400,000 American readers, it can charge \$5 for the book. In this case, revenue is \$3 million in Australia and \$2 million in the United States, for a total of \$5 million. Profit is then \$3 million, which is substantially greater than the \$1 million the company could earn charging the same \$30 price to all customers. Not surprisingly, Readalot chooses to follow this strategy of price discrimination.

Although the story of Readalot Publishing is hypothetical, it describes accurately the business practice of many publishing companies. Textbooks, for example, are often sold at a lower price in Europe than in the United States. Even more important is the price differential between hardcover books and paperbacks. When a publisher has a new novel, it initially releases an expensive hardcover edition and later releases a cheaper paperback edition. The difference in price between these two editions far exceeds the difference in printing costs. The publisher's goal is just as in our example. By selling the hardcover to die-hard fans and the paperback to less enthusiastic readers, the publisher price discriminates and raises its profit.

THE MORAL OF THE STORY

Like any parable, the story of Readalot Publishing is stylized. Yet, also like any parable, it teaches some important and general lessons. In this case, there are three lessons to be learned about price discrimination.

The first and most obvious lesson is that price discrimination is a rational strategy for a profit-maximizing monopolist. In other words, by charging different prices to different customers, a monopolist can increase its profit. In essence, a price-discriminating monopolist charges each customer a price closer to his or her willingness to pay than is possible with a single price.

The second lesson is that price discrimination requires the ability to separate customers according to their willingness to pay. In our example, customers were separated geographically. But sometimes monopolists choose other differences, such as age or income, to distinguish among customers.

A corollary to this second lesson is that certain market forces can prevent firms from price discriminating. In particular, one such force is *arbitrage*, the process of buying a good in one market at a low price and selling it in another market at a higher price in order to profit from the price difference. In our example, suppose that Australian bookstores could buy the book in the United States and resell it to Australian readers. This arbitrage would prevent Readalot from price discriminating because no Australian would buy the book at the higher price.

The third lesson from our parable is perhaps the most surprising: Price discrimination can raise economic welfare. Recall that a deadweight loss arises when Readalot charges a single \$30 price, because the 400,000 less enthusiastic readers do not end up with the book, even though they value it at more than its marginal cost of production. By contrast, when Readalot price discriminates, all readers end up with the book, and the outcome is efficient. Thus, price discrimination can eliminate the inefficiency inherent in monopoly pricing.

Note that the increase in welfare from price discrimination shows up as higher producer surplus rather than higher consumer surplus. In our example, consumers

are no better off for having bought the book: The price they pay exactly equals the value they place on the book, so they receive no consumer surplus. The entire increase in total surplus from price discrimination accrues to Readalot Publishing in the form of higher profit.

THE ANALYTICS OF PRICE DISCRIMINATION

Let's consider a bit more formally how price discrimination affects economic welfare. We begin by assuming that the monopolist can price discriminate perfectly. *Perfect price discrimination* describes a situation in which the monopolist knows exactly the willingness to pay of each customer and can charge each customer a different price. In this case, the monopolist charges each customer exactly his willingness to pay, and the monopolist gets the entire surplus in every transaction.

Figure 15-10 shows producer and consumer surplus with and without price discrimination. Without price discrimination, the firm charges a single price above marginal cost, as shown in panel (a). Because some potential customers who value the good at more than marginal cost do not buy it at this high price, the monopoly causes a deadweight loss. Yet when a firm can perfectly price discriminate, as shown in panel (b), each customer who values the good at more than marginal cost buys the good and is charged his willingness to pay. All mutually beneficial trades take place, there is no deadweight loss, and the entire surplus derived from the market goes to the monopoly producer in the form of profit.

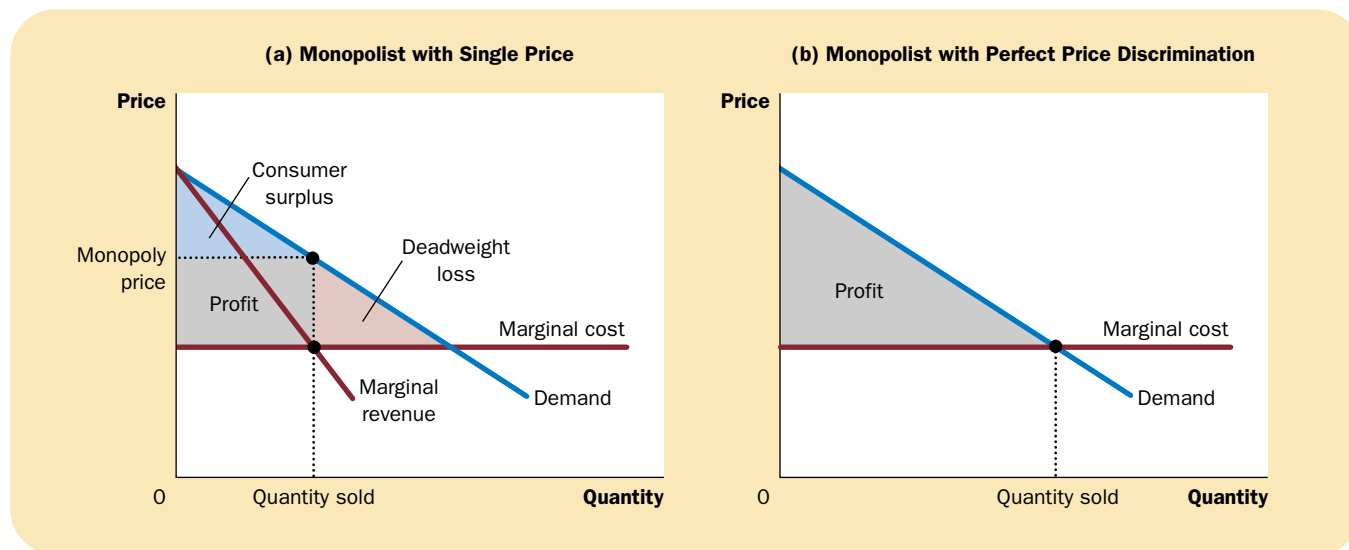


Figure 15-10

WELFARE WITH AND WITHOUT PRICE DISCRIMINATION. Panel (a) shows a monopolist that charges the same price to all customers. Total surplus in this market equals the sum of profit (producer surplus) and consumer surplus. Panel (b) shows a monopolist that can perfectly price discriminate. Because consumer surplus equals zero, total surplus now equals the firm's profit. Comparing these two panels, you can see that perfect price discrimination raises profit, raises total surplus, and lowers consumer surplus.

In reality, of course, price discrimination is not perfect. Customers do not walk into stores with signs displaying their willingness to pay. Instead, firms price discriminate by dividing customers into groups: young versus old, weekday versus weekend shoppers, Americans versus Australians, and so on. Unlike those in our parable of Readalot Publishing, customers within each group differ in their willingness to pay for the product, making perfect price discrimination impossible.

How does this imperfect price discrimination affect welfare? The analysis of these pricing schemes is quite complicated, and it turns out that there is no general answer to this question. Compared to the monopoly outcome with a single price, imperfect price discrimination can raise, lower, or leave unchanged total surplus in a market. The only certain conclusion is that price discrimination raises the monopoly's profit—otherwise the firm would choose to charge all customers the same price.

EXAMPLES OF PRICE DISCRIMINATION

Firms in our economy use various business strategies aimed at charging different prices to different customers. Now that we understand the economics of price discrimination, let's consider some examples.

Movie Tickets Many movie theaters charge a lower price for children and senior citizens than for other patrons. This fact is hard to explain in a competitive market. In a competitive market, price equals marginal cost, and the marginal cost of providing a seat for a child or senior citizen is the same as the marginal cost of providing a seat for anyone else. Yet this fact is easily explained if movie theaters have some local monopoly power and if children and senior citizens have a lower



"Would it bother you to hear how little I paid for this flight?"

IN THE NEWS*The Best Monopolist*

WHAT ORGANIZATION IN OUR ECONOMY IS most successful at exerting market power and keeping prices away from their competitive levels? Economist Robert Barro reports on the first (and only) annual competition to find the most successful monopoly.

Let's Play Monopoly

BY ROBERT J. BARRO

It's almost the end of summer and time for the first annual contest to choose the best operating monopoly in America. The contestants, selected by a panel of Harvard economists, are as follows:

1. The U.S. Postal Service
2. OPEC [Organization of Petroleum Exporting Countries]
3. Almost any cable TV company
4. The Ivy League universities (for administering financial aid to students)
5. The NCAA [National Collegiate Athletic Association] (for administering payments to student-athletes) . . .

Each contestant exhibits fine monopolistic characteristics and is worthy of serious consideration for the award. The U.S. Postal Service claims to be the longest-running monopoly in America and has the distinction of having its control over First Class mail prescribed (perhaps) by the Constitution. The monopoly has preserved large flows of revenues and high wage rates despite studies showing that private companies could carry the mail more efficiently at much lower cost.

On the other hand, the position of the Postal Service has been eroded: first, by successful competition on package delivery; second, by the recent entry of express delivery services; and third, and potentially most damaging, by the introduction of the fax machine. Since faxes are bound to supplant a substantial fraction of First Class letters, the failure to get Congress to classify a fax as First Class mail and, hence, the exclusive domain of the post office shows a remarkable loss of political muscle. Thus, despite past glories, it is hard to be sanguine about the long-term prospects of the post office as a flourishing monopoly.

OPEC was impressive in generating billions of dollars for its members from 1973 to the early 1980s. To understand the functioning of this cartel it is important to sort out the good guys from the bad guys.

The good guys, like Saudi Arabia and Kuwait, are the ones who have typically held oil production below capacity and thereby kept prices above the competitive level. The bad guys, like Libya and Iraq (when Iraq was allowed to produce oil),



IS THE NCAA THE BEST MONOPOLIST?

are the ones who have produced as much as they could and thereby kept prices low.

The good guys were responsible for the vast expansion of oil revenues during the blissful period after 1973. (Hence, they were responsible for the considerable difficulties endured by oil consumers.) But, unfortunately, these countries could not keep the other OPEC members in line and were also unable to exclude new producers or prevent conservation by consumers. Thus, oil prices plummeted in 1986. . . .

In any event, it is unclear that OPEC qualifies for the contest: It is not really American, and its members would

probably be arrested for price-fixing if they ever held an official meeting in America.

Most cable TV companies have government-issued licenses that keep competitors out. Thus, this business supports the hypothesis (offered, I think, by George Stigler) that private monopolies are not sustainable for long unless they have the weight of government behind them.

The rapid escalation of prices and the limitations on services seem, however, to be getting customers and their congressional representatives progressively more annoyed. Thus, it would not be surprising if legislative action leads soon to a deterioration of the cable companies' monopoly power. . . . This fear about the future diminishes the claim of this otherwise worthy contestant for the first annual prize.

Officials of Ivy League universities have been able to meet in semi-public forums to set rules that determine prices of admission (tuition less financial aid) as a function of applicant characteristics, especially financial resources. In some cases, the schools pooled information to agree in advance on the right price to charge a specific customer. Airlines and other industries that wish to price discriminate can only dream about this kind of setup.

Moreover, the universities have more or less successfully applied a high moral tone to the process: Rich applicants—especially smart rich applicants—are charged more than the competitive price for schooling in order to subsidize the smart poor, but it is unclear why this

subsidy should come from the smart rich rather than from taxpayers in general.

In any event, the universities' enviable cartel position has been damaged by the unenlightened Justice Department, which argued that the price-setting meetings were a violation of antitrust laws. Since most of the universities involved have agreed to stop these practices, it may be that future prices for private higher education will come closer to being competitively determined. . . .

The final contestant, the NCAA, has been remarkably successful in holding down "salaries" paid to college athletes. It would be one thing merely to collude to determine price ceilings (for example, to restrict payments so that they not exceed tuition plus room and board and some minor additional amount), but the NCAA has also managed to monopolize all the moral arguments.

Consider a poor ghetto resident who can play basketball well, but not well enough to make it to the NBA. If there were no NCAA, this player might be able legitimately to accumulate a significant amount of cash during a four-year career. But the NCAA ensures that the player will remain poor after four years and, moreover, has convinced most observers that it would be morally wrong for the college to pay the player a competitively determined wage for his or her services.

For many economists, this interference with competition—in a setting that has no obvious reasons for market failure—is itself morally repugnant. But the outrage is compounded here because the transfer is clearly from poor ghetto

residents to rich colleges. Compare the situation of contestant number 4, the Ivy League universities, in which the transfer from rich to poor students can readily be supported on Robin Hood grounds.

The NCAA has the much more difficult task of defending a policy that prevents many poor individuals from earning money. Incredibly, this defense has been so successful that it has even allowed the organization to maintain the moral high ground. When the NCAA maintains its cartel by punishing schools that violate the rules (by paying too much), almost no one doubts that the evil entities are the schools or people who paid the athletes, rather than the cartel enforcers who prevented the athletes from getting paid. Given this extraordinary balancing act, the decision of the panelists was straightforward and the NCAA is the clear and deserving winner of the first annual prize for best monopoly in America.

The panel of economists also considered briefly an award for the least efficient monopoly in America. This choice was, however, too easy. It goes to the American Economic Association, which has been a dismal failure at establishing licensing requirements or other restrictions on entry into the economics profession. It is a sad state of affairs when almost anyone can assume the title of economist.

SOURCE: *The Wall Street Journal*, August 27, 1991, p. A12.

willingness to pay for a ticket. In this case, movie theaters raise their profit by price discriminating.

Airline Prices Seats on airplanes are sold at many different prices. Most airlines charge a lower price for a round-trip ticket between two cities if the traveler stays over a Saturday night. At first this seems odd. Why should it matter to the airline whether a passenger stays over a Saturday night? The reason is that this rule provides a way to separate business travelers and personal travelers. A passenger on a business trip has a high willingness to pay and, most likely, does not want to stay over a Saturday night. By contrast, a passenger traveling for personal reasons has a lower willingness to pay and is more likely to be willing to stay over a Saturday night. Thus, the airlines can successfully price discriminate by charging a lower price for passengers who stay over a Saturday night.

Discount Coupons Many companies offer discount coupons to the public in newspapers and magazines. A buyer simply has to clip out the coupon in order to get \$0.50 off his next purchase. Why do companies offer these coupons? Why don't they just cut the price of the product by \$0.50?

The answer is that coupons allow companies to price discriminate. Companies know that not all customers are willing to spend the time to clip out coupons. Moreover, the willingness to clip coupons is related to the customer's willingness to pay for the good. A rich and busy executive is unlikely to spend her time clipping discount coupons out of the newspaper, and she is probably willing to pay a higher price for many goods. A person who is unemployed is more likely to clip coupons and has a lower willingness to pay. Thus, by charging a lower price only to those customers who clip coupons, firms can successfully price discriminate.

Financial Aid Many colleges and universities give financial aid to needy students. One can view this policy as a type of price discrimination. Wealthy students have greater financial resources and, therefore, a higher willingness to pay than needy students. By charging high tuition and selectively offering financial aid, schools in effect charge prices to customers based on the value they place on going to that school. This behavior is similar to that of any price-discriminating monopolist.

Quantity Discounts So far in our examples of price discrimination, the monopolist charges different prices to different customers. Sometimes, however, monopolists price discriminate by charging different prices to the same customer for different units that the customer buys. For example, many firms offer lower prices to customers who buy large quantities. A bakery might charge \$0.50 for each donut, but \$5 for a dozen. This is a form of price discrimination because the customer pays a higher price for the first unit bought than for the twelfth. Quantity discounts are often a successful way of price discriminating because a customer's willingness to pay for an additional unit declines as the customer buys more units.

QUICK QUIZ: Give two examples of price discrimination. ♦ How does perfect price discrimination affect consumer surplus, producer surplus, and total surplus?

CONCLUSION: THE PREVALENCE OF MONOPOLY

This chapter has discussed the behavior of firms that have control over the prices they charge. We have seen that because monopolists produce less than the socially efficient quantity and charge prices above marginal cost, they cause deadweight losses. These inefficiencies can be mitigated through prudent public policies or, in some cases, through price discrimination by the monopolist.

How prevalent are the problems of monopoly? There are two answers to this question.

In one sense, monopolies are common. Most firms have some control over the prices they charge. They are not forced to charge the market price for their goods, because their goods are not exactly the same as those offered by other firms. A Ford Taurus is not the same as a Toyota Camry. Ben and Jerry's ice cream is not the same as Breyer's. Each of these goods has a downward-sloping demand curve, which gives each producer some degree of monopoly power.

Yet firms with substantial monopoly power are quite rare. Few goods are truly unique. Most have substitutes that, even if not exactly the same, are very similar. Ben and Jerry can raise the price of their ice cream a little without losing all their sales; but if they raise it very much, sales will fall substantially.

In the end, monopoly power is a matter of degree. It is true that many firms have some monopoly power. It is also true that their monopoly power is usually quite limited. In these cases, we will not go far wrong assuming that firms operate in competitive markets, even if that is not precisely the case.

Summary

- ◆ A monopoly is a firm that is the sole seller in its market. A monopoly arises when a single firm owns a key resource, when the government gives a firm the exclusive right to produce a good, or when a single firm can supply the entire market at a smaller cost than many firms could.
- ◆ Because a monopoly is the sole producer in its market, it faces a downward-sloping demand curve for its product. When a monopoly increases production by 1 unit, it causes the price of its good to fall, which reduces the amount of revenue earned on all units produced. As a result, a monopoly's marginal revenue is always below the price of its good.
- ◆ Like a competitive firm, a monopoly firm maximizes profit by producing the quantity at which marginal revenue equals marginal cost. The monopoly then chooses the price at which that quantity is demanded. Unlike a competitive firm, a monopoly firm's price exceeds its marginal revenue, so its price exceeds marginal cost.
- ◆ A monopolist's profit-maximizing level of output is below the level that maximizes the sum of consumer and producer surplus. That is, when the monopoly charges a price above marginal cost, some consumers who value the good more than its cost of production do not buy it. As a result, monopoly causes deadweight losses similar to the deadweight losses caused by taxes.
- ◆ Policymakers can respond to the inefficiency of monopoly behavior in four ways. They can use the antitrust laws to try to make the industry more competitive. They can regulate the prices that the monopoly charges. They can turn the monopolist into a government-run enterprise. Or, if the market failure is deemed small compared to the inevitable imperfections of policies, they can do nothing at all.
- ◆ Monopolists often can raise their profits by charging different prices for the same good based on a buyer's willingness to pay. This practice of price discrimination can raise economic welfare by getting the good to some

consumers who otherwise would not buy it. In the extreme case of perfect price discrimination, the deadweight losses of monopoly are completely

eliminated. More generally, when price discrimination is imperfect, it can either raise or lower welfare compared to the outcome with a single monopoly price.

Key Concepts

monopoly, p. 316

natural monopoly, p. 318

price discrimination, p. 336

Questions for Review

1. Give an example of a government-created monopoly. Is creating this monopoly necessarily bad public policy? Explain.
2. Define natural monopoly. What does the size of a market have to do with whether an industry is a natural monopoly?
3. Why is a monopolist's marginal revenue less than the price of its good? Can marginal revenue ever be negative? Explain.
4. Draw the demand, marginal-revenue, and marginal-cost curves for a monopolist. Show the profit-maximizing level of output. Show the profit-maximizing price.
5. In your diagram from the previous question, show the level of output that maximizes total surplus. Show the deadweight loss from the monopoly. Explain your answer.
6. What gives the government the power to regulate mergers between firms? From the standpoint of the welfare of society, give a good reason and a bad reason that two firms might want to merge.
7. Describe the two problems that arise when regulators tell a natural monopoly that it must set a price equal to marginal cost.
8. Give two examples of price discrimination. In each case, explain why the monopolist chooses to follow this business strategy.

Problems and Applications

1. A publisher faces the following demand schedule for the next novel by one of its popular authors:

PRICE	QUANTITY DEMANDED
\$100	0
90	100,000
80	200,000
70	300,000
60	400,000
50	500,000
40	600,000
30	700,000
20	800,000
10	900,000
0	1,000,000

The author is paid \$2 million to write the book, and the marginal cost of publishing the book is a constant \$10 per book.

- a. Compute total revenue, total cost, and profit at each quantity. What quantity would a profit-maximizing publisher choose? What price would it charge?
- b. Compute marginal revenue. (Recall that $MR = \Delta TR / \Delta Q$.) How does marginal revenue compare to the price? Explain.
- c. Graph the marginal-revenue, marginal-cost, and demand curves. At what quantity do the marginal-revenue and marginal-cost curves cross? What does this signify?
- d. In your graph, shade in the deadweight loss. Explain in words what this means.

- e. If the author were paid \$3 million instead of \$2 million to write the book, how would this affect the publisher's decision regarding the price to charge? Explain.
 - f. Suppose the publisher were not profit-maximizing but were concerned with maximizing economic efficiency. What price would it charge for the book? How much profit would it make at this price?
2. Suppose that a natural monopolist were required by law to charge average total cost. On a diagram, label the price charged and the deadweight loss to society relative to marginal-cost pricing.
 3. Consider the delivery of mail. In general, what is the shape of the average-total-cost curve? How might the shape differ between isolated rural areas and densely populated urban areas? How might the shape have changed over time? Explain.
 4. Suppose the Clean Springs Water Company has a monopoly on bottled water sales in California. If the price of tap water increases, what is the change in Clean Springs' profit-maximizing levels of output, price, and profit? Explain in words and with a graph.
 5. A small town is served by many competing supermarkets, which have constant marginal cost.
 - a. Using a diagram of the market for groceries, show the consumer surplus, producer surplus, and total surplus.
 - b. Now suppose that the independent supermarkets combine into one chain. Using a new diagram, show the new consumer surplus, producer surplus, and total surplus. Relative to the competitive market, what is the transfer from consumers to producers? What is the deadweight loss?
 6. Johnny Rockabilly has just finished recording his latest CD. His record company's marketing department determines that the demand for the CD is as follows:

PRICE	NUMBER OF CDs
\$24	10,000
22	20,000
20	30,000
18	40,000
16	50,000
14	60,000

The company can produce the CD with no fixed cost and a variable cost of \$5 per CD.

- a. Find total revenue for quantity equal to 10,000, 20,000, and so on. What is the marginal revenue for each 10,000 increase in the quantity sold?
 - b. What quantity of CDs would maximize profit? What would the price be? What would the profit be?
 - c. If you were Johnny's agent, what recording fee would you advise Johnny to demand from the record company? Why?
7. In 1969 the government charged IBM with monopolizing the computer market. The government argued (correctly) that a large share of all mainframe computers sold in the United States were produced by IBM. IBM argued (correctly) that a much smaller share of the market for *all* types of computers consisted of IBM products. Based on these facts, do you think that the government should have brought suit against IBM for violating the antitrust laws? Explain.
 8. A company is considering building a bridge across a river. The bridge would cost \$2 million to build and nothing to maintain. The following table shows the company's anticipated demand over the lifetime of the bridge:

PRICE (PER CROSSING)	NUMBER OF CROSSINGS (IN THOUSANDS)
\$8	0
7	100
6	200
5	300
4	400
3	500
2	600
1	700
0	800

- a. If the company were to build the bridge, what would be its profit-maximizing price? Would that be the efficient level of output? Why or why not?
 - b. If the company is interested in maximizing profit, should it build the bridge? What would be its profit or loss?
 - c. If the government were to build the bridge, what price should it charge?
 - d. Should the government build the bridge? Explain.
9. The Placebo Drug Company holds a patent on one of its discoveries.

- a. Assuming that the production of the drug involves rising marginal cost, draw a diagram to illustrate Placebo's profit-maximizing price and quantity. Also show Placebo's profits.
 - b. Now suppose that the government imposes a tax on each bottle of the drug produced. On a new diagram, illustrate Placebo's new price and quantity. How does each compare to your answer in part (a)?
 - c. Although it is not easy to see in your diagrams, the tax reduces Placebo's profit. Explain why this must be true.
 - d. Instead of the tax per bottle, suppose that the government imposes a tax on Placebo of \$10,000 regardless of how many bottles are produced. How does this tax affect Placebo's price, quantity, and profits? Explain.
10. Larry, Curly, and Moe run the only saloon in town. Larry wants to sell as many drinks as possible without losing money. Curly wants the saloon to bring in as much revenue as possible. Moe wants to make the largest possible profits. Using a single diagram of the saloon's demand curve and its cost curves, show the price and quantity combinations favored by each of the three partners. Explain.
 11. For many years AT&T was a regulated monopoly, providing both local and long-distance telephone service.
 - a. Explain why long-distance phone service was originally a natural monopoly.
 - b. Over the past two decades, many companies have launched communication satellites, each of which can transmit a limited number of calls. How did the growing role of satellites change the cost structure of long-distance phone service?

After a lengthy legal battle with the government, AT&T agreed to compete with other companies in the long-distance market. It also agreed to spin off its local phone service into the "Baby Bells," which remain highly regulated.

 - c. Why might it be efficient to have competition in long-distance phone service and regulated monopolies in local phone service?
 12. The Best Computer Company just developed a new computer chip, on which it immediately acquires a patent.
 - a. Draw a diagram that shows the consumer surplus, producer surplus, and total surplus in the market for this new chip.
 - b. What happens to these three measures of surplus if the firm can perfectly price discriminate? What is the change in deadweight loss? What transfers occur?
 13. Explain why a monopolist will always produce a quantity at which the demand curve is elastic. (Hint: If demand is inelastic and the firm raises its price, what happens to total revenue and total costs?)
 14. The "Big Three" American car companies are GM, Ford, and Chrysler. If these were the only car companies in the world, they would have much more monopoly power. What action could the U.S. government take to create monopoly power for these companies? (Hint: The government took such an action in the 1980s.)
 15. Singer Whitney Houston has a monopoly over a scarce resource: herself. She is the only person who can produce a Whitney Houston concert. Does this fact imply that the government should regulate the prices of her concerts? Why or why not?
 16. Many schemes for price discriminating involve some cost. For example, discount coupons take up time and resources from both the buyer and the seller. This question considers the implications of costly price discrimination. To keep things simple, let's assume that our monopolist's production costs are simply proportional to output, so that average total cost and marginal cost are constant and equal to each other.
 - a. Draw the cost, demand, and marginal-revenue curves for the monopolist. Show the price the monopolist would charge without price discrimination.
 - b. In your diagram, mark the area equal to the monopolist's profit and call it X . Mark the area equal to consumer surplus and call it Y . Mark the area equal to the deadweight loss and call it Z .
 - c. Now suppose that the monopolist can perfectly price discriminate. What is the monopolist's profit? (Give your answer in terms of X , Y , and Z .)
 - d. What is the change in the monopolist's profit from price discrimination? What is the change in total surplus from price discrimination? Which change is larger? Explain. (Give your answer in terms of X , Y , and Z .)
 - e. Now suppose that there is some cost of price discrimination. To model this cost, let's assume that the monopolist has to pay a fixed cost C in order to price discriminate. How would a monopolist make the decision whether to pay this fixed cost? (Give your answer in terms of X , Y , Z , and C .)

- f. How would a benevolent social planner, who cares about total surplus, decide whether the monopolist should price discriminate? (Give your answer in terms of X , Y , Z , and C .)
- g. Compare your answers to parts (e) and (f). How does the monopolist's incentive to price

discriminate differ from the social planner's? Is it possible that the monopolist will price discriminate even though it is not socially desirable?

16



OLIGOPOLY

If you go to a store to buy tennis balls, it is likely that you will come home with one of four brands: Wilson, Penn, Dunlop, or Spalding. These four companies make almost all of the tennis balls sold in the United States. Together these firms determine the quantity of tennis balls produced and, given the market demand curve, the price at which tennis balls are sold.

How can we describe the market for tennis balls? The previous two chapters discussed two types of market structure. In a competitive market, each firm is so small compared to the market that it cannot influence the price of its product and, therefore, takes the price as given by market conditions. In a monopolized market, a single firm supplies the entire market for a good, and that firm can choose any price and quantity on the market demand curve.

The market for tennis balls fits neither the competitive nor the monopoly model. Competition and monopoly are extreme forms of market structure. Competition occurs when there are many firms in a market offering essentially identical products; monopoly occurs when there is only one firm in a market. It is

**IN THIS CHAPTER
YOU WILL . . .**

See what market structures lie between monopoly and competition

Examine what outcomes are possible when a market is an oligopoly

Learn about the prisoners' dilemma and how it applies to oligopoly and other issues

Consider how the antitrust laws try to foster competition in oligopolistic markets

natural to start the study of industrial organization with these polar cases, for they are the easiest cases to understand. Yet many industries, including the tennis ball industry, fall somewhere between these two extremes. Firms in these industries have competitors but, at the same time, do not face so much competition that they are price takers. Economists call this situation *imperfect competition*.

In this chapter we discuss the types of imperfect competition and examine a particular type called *oligopoly*. The essence of an oligopolistic market is that there are only a few sellers. As a result, the actions of any one seller in the market can have a large impact on the profits of all the other sellers. That is, oligopolistic firms are interdependent in a way that competitive firms are not. Our goal in this chapter is to see how this interdependence shapes the firms' behavior and what problems it raises for public policy.

BETWEEN MONOPOLY AND PERFECT COMPETITION

The previous two chapters analyzed markets with many competitive firms and markets with a single monopoly firm. In Chapter 14, we saw that the price in a perfectly competitive market always equals the marginal cost of production. We also saw that, in the long run, entry and exit drive economic profit to zero, so the price also equals average total cost. In Chapter 15, we saw how firms with market power can use that power to keep prices above marginal cost, leading to a positive economic profit for the firm and a deadweight loss for society.

The cases of perfect competition and monopoly illustrate some important ideas about how markets work. Most markets in the economy, however, include elements of both these cases and, therefore, are not completely described by either of them. The typical firm in the economy faces competition, but the competition is not so rigorous as to make the firm exactly described by the price-taking firm analyzed in Chapter 14. The typical firm also has some degree of market power, but its market power is not so great that the firm can be exactly described by the monopoly firm analyzed in Chapter 15. In other words, the typical firm in our economy is imperfectly competitive.

oligopoly

a market structure in which only a few sellers offer similar or identical products

monopolistic competition

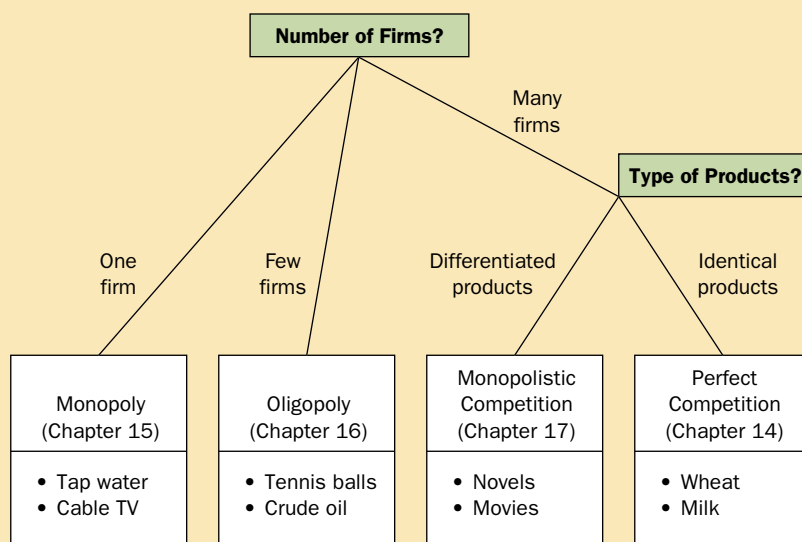
a market structure in which many firms sell products that are similar but not identical

There are two types of imperfectly competitive markets. An **oligopoly** is a market with only a few sellers, each offering a product similar or identical to the others. One example is the market for tennis balls. Another is the world market for crude oil: A few countries in the Middle East control much of the world's oil reserves. **Monopolistic competition** describes a market structure in which there are many firms selling products that are similar but not identical. Examples include the markets for novels, movies, CDs, and computer games. In a monopolistically competitive market, each firm has a monopoly over the product it makes, but many other firms make similar products that compete for the same customers.

Figure 16-1 summarizes the four types of market structure. The first question to ask about any market is how many firms there are. If there is only one firm, the market is a monopoly. If there are only a few firms, the market is an oligopoly. If there are many firms, we need to ask another question: Do the firms sell identical or differentiated products? If the many firms sell differentiated products, the market is monopolistically competitive. If the many firms sell identical products, the market is perfectly competitive.

Figure 16-1

THE FOUR TYPES OF MARKET STRUCTURE. Economists who study industrial organization divide markets into four types—monopoly, oligopoly, monopolistic competition, and perfect competition.



Reality, of course, is never as clear-cut as theory. In some cases, you may find it hard to decide what structure best describes a market. There is, for instance, no magic number that separates “few” from “many” when counting the number of firms. (Do the approximately dozen companies that now sell cars in the United States make this market an oligopoly or more competitive? The answer is open to debate.) Similarly, there is no sure way to determine when products are differentiated and when they are identical. (Are different brands of milk really the same? Again, the answer is debatable.) When analyzing actual markets, economists have to keep in mind the lessons learned from studying all types of market structure and then apply each lesson as it seems appropriate.

Now that we understand how economists define the various types of market structure, we can continue our analysis of them. In the next chapter we analyze monopolistic competition. In this chapter we examine oligopoly.

QUICK QUIZ: Define *oligopoly* and *monopolistic competition* and give an example of each.

MARKETS WITH ONLY A FEW SELLERS

Because an oligopolistic market has only a small group of sellers, a key feature of oligopoly is the tension between cooperation and self-interest. The group of oligopolists is best off cooperating and acting like a monopolist—producing a

small quantity of output and charging a price above marginal cost. Yet because each oligopolist cares about only its own profit, there are powerful incentives at work that hinder a group of firms from maintaining the monopoly outcome.

A DUOPOLY EXAMPLE

To understand the behavior of oligopolies, let's consider an oligopoly with only two members, called a *duopoly*. Duopoly is the simplest type of oligopoly. Oligopolies with three or more members face the same problems as oligopolies with only two members, so we do not lose much by starting with the case of duopoly.

Imagine a town in which only two residents—Jack and Jill—own wells that produce water safe for drinking. Each Saturday, Jack and Jill decide how many gallons of water to pump, bring the water to town, and sell it for whatever price the market will bear. To keep things simple, suppose that Jack and Jill can pump as much water as they want without cost. That is, the marginal cost of water equals zero.

Table 16-1 shows the town's demand schedule for water. The first column shows the total quantity demanded, and the second column shows the price. If the two well owners sell a total of 10 gallons of water, water goes for \$110 a gallon. If they sell a total of 20 gallons, the price falls to \$100 a gallon. And so on. If you graphed these two columns of numbers, you would get a standard downward-sloping demand curve.

The last column in Table 16-1 shows the total revenue from the sale of water. It equals the quantity sold times the price. Because there is no cost to pumping water, the total revenue of the two producers equals their total profit.

Let's now consider how the organization of the town's water industry affects the price of water and the quantity of water sold.

Table 16-1

THE DEMAND SCHEDULE
FOR WATER

QUANTITY (IN GALLONS)	PRICE	TOTAL REVENUE (AND TOTAL PROFIT)
0	\$120	\$ 0
10	110	1,100
20	100	2,000
30	90	2,700
40	80	3,200
50	70	3,500
60	60	3,600
70	50	3,500
80	40	3,200
90	30	2,700
100	20	2,000
110	10	1,100
120	0	0

COMPETITION, MONOPOLIES, AND CARTELS

Before considering the price and quantity of water that would result from the duopoly of Jack and Jill, let's discuss briefly the two market structures we already understand: competition and monopoly.

Consider first what would happen if the market for water were perfectly competitive. In a competitive market, the production decisions of each firm drive price equal to marginal cost. In the market for water, marginal cost is zero. Thus, under competition, the equilibrium price of water would be zero, and the equilibrium quantity would be 120 gallons. The price of water would reflect the cost of producing it, and the efficient quantity of water would be produced and consumed.

Now consider how a monopoly would behave. Table 16-1 shows that total profit is maximized at a quantity of 60 gallons and a price of \$60 a gallon. A profit-maximizing monopolist, therefore, would produce this quantity and charge this price. As is standard for monopolies, price would exceed marginal cost. The result would be inefficient, for the quantity of water produced and consumed would fall short of the socially efficient level of 120 gallons.

What outcome should we expect from our duopolists? One possibility is that Jack and Jill get together and agree on the quantity of water to produce and the price to charge for it. Such an agreement among firms over production and price is called **collusion**, and the group of firms acting in unison is called a **cartel**. Once a cartel is formed, the market is in effect served by a monopoly, and we can apply our analysis from Chapter 15. That is, if Jack and Jill were to collude, they would agree on the monopoly outcome because that outcome maximizes the total profit that the producers can get from the market. Our two producers would produce a total of 60 gallons, which would be sold at a price of \$60 a gallon. Once again, price exceeds marginal cost, and the outcome is socially inefficient.

A cartel must agree not only on the total level of production but also on the amount produced by each member. In our case, Jack and Jill must agree how to split between themselves the monopoly production of 60 gallons. Each member of the cartel will want a larger share of the market because a larger market share means larger profit. If Jack and Jill agreed to split the market equally, each would produce 30 gallons, the price would be \$60 a gallon, and each would get a profit of \$1,800.

collusion

an agreement among firms in a market about quantities to produce or prices to charge

cartel

a group of firms acting in unison

THE EQUILIBRIUM FOR AN OLIGOPOLY

Although oligopolists would like to form cartels and earn monopoly profits, often that is not possible. As we discuss later in this chapter, antitrust laws prohibit explicit agreements among oligopolists as a matter of public policy. In addition, squabbling among cartel members over how to divide the profit in the market sometimes makes agreement among them impossible. Let's therefore consider what happens if Jack and Jill decide separately how much water to produce.

At first, one might expect Jack and Jill to reach the monopoly outcome on their own, for this outcome maximizes their joint profit. In the absence of a binding agreement, however, the monopoly outcome is unlikely. To see why, imagine that Jack expects Jill to produce only 30 gallons (half of the monopoly quantity). Jack would reason as follows:

IN THE NEWS

Modern Pirates

CARTELS ARE RARE, IN PART BECAUSE THE antitrust laws make them illegal. As the following article describes, however, ocean shipping firms enjoy an unusual exemption from these laws and, as a result, charge higher prices than they otherwise would.

As U.S. Trade Grows, Shipping Cartels Get a Bit More Scrutiny

BY ANNA WILDE MATTHEWS

RUTHERFORD, N.J.—Every two weeks, in an unobtrusive office building here, about 20 shipping-line managers gather for their usual meeting. They sit around a long conference table, exchange small talk over bagels and coffee and then begin discussing what they will charge to move cargo across the Atlantic Ocean.

All very routine, except for one detail: They don't work for the same company. Each represents a different shipping line, supposedly competing for business. Under U.S. antitrust law, most people doing this would end up in court.

But shipping isn't like other businesses. Many of the world's big shipping lines, from Sea-Land Service Inc. of the U.S. to A. P. Moller/Maersk Line of Denmark, are members of a little-noticed cartel that for many decades has set rates on tens of billions of dollars of cargo.

Most U.S. consumer goods exported or imported by sea are affected

to some degree. The cartel—really a series of cartels, one for each major shipping route—can tell importers and exporters when shipping contracts start and when they end. They can favor one port over another, enough to swing badly needed trade away from an entire city. And because the shipping industry has an antitrust exemption from Congress, all of this is legal.

"This is one of the last legalized price-setting arrangements in existence," says Robert Litan, a former Justice Department antitrust official. Airlines and banks couldn't do this, he says, "but if you're an ocean shipping line, there's nothing to stop you from price fixing."

You could call them the OPEC of shipping, though not quite as powerful because they can't keep members from building too many ships. To get more business, some of the shipping cartels' own members undercut cartel rates or make special deals with big customers. They also face the emergence of new competitors, which are keeping rates down in some markets.

Nonetheless, the industry is playing a bigger role now in the U.S. economy as American companies plunge more deeply into world trade. Exports over the seas have jumped 26% in the past two years and 50% since the start of the decade.

For consumers, the impact is hard to measure. Transportation costs make up 5% to 10% of the price of most goods, and increases in shipping rates are usually passed on to consumers. A limited 1993 survey by the Agriculture Department, examining \$5 billion of U.S. farm exports, concluded that the cartels were raising ocean shipping rates as much as 18%. A different report, by the Federal Trade Commission in 1995, found that when shipping lines broke

free of cartel rates, contract prices were about 19% lower.

"The cartels' whole makeup is anti-consumer," says John Taylor, a transportation professor at Wayne State University in Detroit. "They're designed to keep prices up."

Some moves are afoot to change all this. The U.S. Senate is considering a bill that, for the first time in a decade, would weaken the cartels, by reducing their power to police their members. The bill, sponsored by Sen. Kay Bailey Hutchison of Texas, has the support of some other high-ranking Republicans, including Majority Leader Trent Lott. . . .

For eight decades, shipping cartels have been protected by Congress under the Shipping Act of 1916, passed at the behest of American shipping customers, who thought cartels would guarantee reliable service. The law was revised significantly only twice, in 1961 and 1984, but both times the industry's antitrust immunity was left intact.

The most recent major review was done in 1991 by a congressional commission. It heard more than 100 witnesses, produced a 250-page report—and offered no conclusions or recommendations. . . .

The real reasons for years of inaction in Congress may be apathy and the lobbying by various groups. Dockside labor, for example, fears that secret contracts would enable ship lines to divert cargo to nonunion workers without the union knowing it. David Butz, a University of Michigan economist who has studied shipping, thinks voters aren't likely to weigh in; the cartels aren't a hot topic. "It's below the radar screen," he says. "Consumers don't realize the impact they have."

SOURCE: *The Wall Street Journal*, October 7, 1997, p. A1.

"I could produce 30 gallons as well. In this case, a total of 60 gallons of water would be sold at a price of \$60 a gallon. My profit would be \$1,800 (30 gallons $\times$ \$60 a gallon). Alternatively, I could produce 40 gallons. In this case, a total of 70 gallons of water would be sold at a price of \$50 a gallon. My profit would be \$2,000 (40 gallons $\times$ \$50 a gallon). Even though total profit in the market would fall, my profit would be higher, because I would have a larger share of the market."

Of course, Jill might reason the same way. If so, Jack and Jill would each bring 40 gallons to town. Total sales would be 80 gallons, and the price would fall to \$40. Thus, if the duopolists individually pursue their own self-interest when deciding how much to produce, they produce a total quantity greater than the monopoly quantity, charge a price lower than the monopoly price, and earn total profit less than the monopoly profit.

Although the logic of self-interest increases the duopoly's output above the monopoly level, it does not push the duopolists to reach the competitive allocation. Consider what happens when each duopolist is producing 40 gallons. The price is \$40, and each duopolist makes a profit of \$1,600. In this case, Jack's self-interested logic leads to a different conclusion:

"Right now, my profit is \$1,600. Suppose I increase my production to 50 gallons. In this case, a total of 90 gallons of water would be sold, and the price would be \$30 a gallon. Then my profit would be only \$1,500. Rather than increasing production and driving down the price, I am better off keeping my production at 40 gallons."

The outcome in which Jack and Jill each produce 40 gallons looks like some sort of equilibrium. In fact, this outcome is called a *Nash equilibrium* (named after economic theorist John Nash). A **Nash equilibrium** is a situation in which economic actors interacting with one another each choose their best strategy given the strategies the others have chosen. In this case, given that Jill is producing 40 gallons, the best strategy for Jack is to produce 40 gallons. Similarly, given that Jack is producing 40 gallons, the best strategy for Jill is to produce 40 gallons. Once they reach this Nash equilibrium, neither Jack nor Jill has an incentive to make a different decision.

This example illustrates the tension between cooperation and self-interest. Oligopolists would be better off cooperating and reaching the monopoly outcome. Yet because they pursue their own self-interest, they do not end up reaching the monopoly outcome and maximizing their joint profit. Each oligopolist is tempted to raise production and capture a larger share of the market. As each of them tries to do this, total production rises, and the price falls.

At the same time, self-interest does not drive the market all the way to the competitive outcome. Like monopolists, oligopolists are aware that increases in the amount they produce reduce the price of their product. Therefore, they stop short of following the competitive firm's rule of producing up to the point where price equals marginal cost.

In summary, *when firms in an oligopoly individually choose production to maximize profit, they produce a quantity of output greater than the level produced by monopoly and less than the level produced by competition. The oligopoly price is less than the monopoly price but greater than the competitive price (which equals marginal cost).*

Nash equilibrium

a situation in which economic actors interacting with one another each choose their best strategy given the strategies that all the other actors have chosen

HOW THE SIZE OF AN OLIGOPOLY AFFECTS THE MARKET OUTCOME

We can use the insights from this analysis of duopoly to discuss how the size of an oligopoly is likely to affect the outcome in a market. Suppose, for instance, that

John and Joan suddenly discover water sources on their property and join Jack and Jill in the water oligopoly. The demand schedule in Table 16-1 remains the same, but now more producers are available to satisfy this demand. How would an increase in the number of sellers from two to four affect the price and quantity of water in the town?

If the sellers of water could form a cartel, they would once again try to maximize total profit by producing the monopoly quantity and charging the monopoly price. Just as when there were only two sellers, the members of the cartel would need to agree on production levels for each member and find some way to enforce the agreement. As the cartel grows larger, however, this outcome is less likely. Reaching and enforcing an agreement becomes more difficult as the size of the group increases.

If the oligopolists do not form a cartel—perhaps because the antitrust laws prohibit it—they must each decide on their own how much water to produce. To see how the increase in the number of sellers affects the outcome, consider the decision facing each seller. At any time, each well owner has the option to raise production by 1 gallon. In making this decision, the well owner weighs two effects:

- ◆ *The output effect:* Because price is above marginal cost, selling 1 more gallon of water at the going price will raise profit.
- ◆ *The price effect:* Raising production will increase the total amount sold, which will lower the price of water and lower the profit on all the other gallons sold.

If the output effect is larger than the price effect, the well owner will increase production. If the price effect is larger than the output effect, the owner will not raise production. (In fact, in this case, it is profitable to reduce production.) Each oligopolist continues to increase production until these two marginal effects exactly balance, taking the other firms' production as given.

Now consider how the number of firms in the industry affects the marginal analysis of each oligopolist. The larger the number of sellers, the less concerned each seller is about its own impact on the market price. That is, as the oligopoly grows in size, the magnitude of the price effect falls. When the oligopoly grows very large, the price effect disappears altogether, leaving only the output effect. In this extreme case, each firm in the oligopoly increases production as long as price is above marginal cost.

We can now see that a large oligopoly is essentially a group of competitive firms. A competitive firm considers only the output effect when deciding how much to produce: Because a competitive firm is a price taker, the price effect is absent. Thus, *as the number of sellers in an oligopoly grows larger, an oligopolistic market looks more and more like a competitive market. The price approaches marginal cost, and the quantity produced approaches the socially efficient level.*

This analysis of oligopoly offers a new perspective on the effects of international trade. Imagine that Toyota and Honda are the only automakers in Japan, Volkswagen and Mercedes-Benz are the only automakers in Germany, and Ford and General Motors are the only automakers in the United States. If these nations prohibited trade in autos, each would have an auto oligopoly with only two members, and the market outcome would likely depart substantially from the competitive ideal. With international trade, however, the car market is a world market, and the oligopoly in this example has six members. Allowing free trade increases

the number of producers from which each consumer can choose, and this increased competition keeps prices closer to marginal cost. Thus, the theory of oligopoly provides another reason, in addition to the theory of comparative advantage discussed in Chapter 3, why all countries can benefit from free trade.

CASE STUDY OPEC AND THE WORLD OIL MARKET

Our story about the town's market for water is fictional, but if we change water to crude oil, and Jack and Jill to Iran and Iraq, the story is quite close to being true. Much of the world's oil is produced by a few countries, mostly in the Middle East. These countries together make up an oligopoly. Their decisions about how much oil to pump are much the same as Jack and Jill's decisions about how much water to pump.

The countries that produce most of the world's oil have formed a cartel, called the Organization of Petroleum Exporting Countries (OPEC). As originally formed in 1960, OPEC included Iran, Iraq, Kuwait, Saudi Arabia, and Venezuela. By 1973, eight other nations had joined: Qatar, Indonesia, Libya, the United Arab Emirates, Algeria, Nigeria, Ecuador, and Gabon. These countries control about three-fourths of the world's oil reserves. Like any cartel, OPEC tries to raise the price of its product through a coordinated reduction in quantity produced. OPEC tries to set production levels for each of the member countries.

The problem that OPEC faces is much the same as the problem that Jack and Jill face in our story. The OPEC countries would like to maintain a high price of oil. But each member of the cartel is tempted to increase production in order to get a larger share of the total profit. OPEC members frequently agree to reduce production but then cheat on their agreements.

OPEC was most successful at maintaining cooperation and high prices in the period from 1973 to 1985. The price of crude oil rose from \$2.64 a barrel in 1972 to \$11.17 in 1974 and then to \$35.10 in 1981. But in the early 1980s member countries began arguing about production levels, and OPEC became ineffective at maintaining cooperation. By 1986 the price of crude oil had fallen back to \$12.52 a barrel.

During the 1990s, the members of OPEC met about twice a year, but the cartel failed to reach and enforce agreement. The members of OPEC made production decisions largely independently of one another, and the world market for oil was fairly competitive. Throughout most of the decade, the price of crude oil, adjusted for overall inflation, remained less than half the level OPEC had achieved in 1981. In 1999, however, cooperation among oil-exporting nations started to pick up (see the accompanying In the News box). Only time will tell how persistent this renewed cooperation proves to be.



OPEC: A NOT VERY COOPERATIVE CARTEL

QUICK QUIZ: If the members of an oligopoly could agree on a total quantity to produce, what quantity would they choose? ♦ If the oligopolists do not act together but instead make production decisions individually, do they produce a total quantity more or less than in your answer to the previous question? Why?

IN THE NEWS*The Oil Cartel
Makes a Comeback*

OPEC FAILED TO KEEP OIL PRICES HIGH during most of the 1990s, but this started to change in 1999.

An Oil Outsider Revives a Cartel

BY AGIS SALPUKAS

The price of crude oil has doubled since early last year. Higher prices for gaso-

line, heating oil, and other products are hitting every consumer's pocketbook.

Is OPEC flexing its muscle again? Not exactly. There's a new cartel in town, and after a shaky start two years ago, its members have achieved—for now, at least—the unity necessary to hold to their production quotas. And that means higher prices.

In a sense, this cartel is simply the 11 members of the Organization of Petroleum Exporting Countries plus two—Mexico and Norway. But the world's oil-producing and exporting nations are wielding power this time around mainly because of a shove not from the Middle East but rather from Mexico—and especially from its persistent energy minister, Luis K. Tellez. . . . Already, the price of

crude oil has more than doubled, to \$23.45 a barrel from \$11 early this year.

Not that the coalition is home free. Prices hit \$24 a barrel last month, but slipped back when traders thought they saw hints of cracks in the cartel's solidarity. After all, if one country breaks ranks, the cartel's tenuous grip on the world market could crumble.

For the moment, though, there seems little easing in the cartel's united front or in rising oil prices.

SOURCE: *The New York Times*, Money & Business Section, October 24, 1999, p. 1.

GAME THEORY AND THE ECONOMICS OF COOPERATION

As we have seen, oligopolies would like to reach the monopoly outcome, but doing so requires cooperation, which at times is difficult to maintain. In this section we look more closely at the problems people face when cooperation is desirable but difficult. To analyze the economics of cooperation, we need to learn a little about game theory.

game theory

the study of how people behave in strategic situations

Game theory is the study of how people behave in strategic situations. By “strategic” we mean a situation in which each person, when deciding what actions to take, must consider how others might respond to that action. Because the number of firms in an oligopolistic market is small, each firm must act strategically. Each firm knows that its profit depends not only on how much it produces but also on how much the other firms produce. In making its production decision, each firm in an oligopoly should consider how its decision might affect the production decisions of all the other firms.

Game theory is not necessary for understanding competitive or monopoly markets. In a competitive market, each firm is so small compared to the market that strategic interactions with other firms are not important. In a monopolized market, strategic interactions are absent because the market has only one firm. But, as we will see, game theory is quite useful for understanding the behavior of oligopolies.

A particularly important “game” is called the **prisoners’ dilemma**. This game provides insight into the difficulty of maintaining cooperation. Many times in life, people fail to cooperate with one another even when cooperation would make them all better off. An oligopoly is just one example. The story of the prisoners’ dilemma contains a general lesson that applies to any group trying to maintain cooperation among its members.

prisoners’ dilemma

a particular “game” between two captured prisoners that illustrates why cooperation is difficult to maintain even when it is mutually beneficial

THE PRISONERS’ DILEMMA

The prisoners’ dilemma is a story about two criminals who have been captured by the police. Let’s call them Bonnie and Clyde. The police have enough evidence to convict Bonnie and Clyde of the minor crime of carrying an unregistered gun, so that each would spend a year in jail. The police also suspect that the two criminals have committed a bank robbery together, but they lack hard evidence to convict them of this major crime. The police question Bonnie and Clyde in separate rooms, and they offer each of them the following deal:

“Right now, we can lock you up for 1 year. If you confess to the bank robbery and implicate your partner, however, we’ll give you immunity and you can go free. Your partner will get 20 years in jail. But if you both confess to the crime, we won’t need your testimony and we can avoid the cost of a trial, so you will each get an intermediate sentence of 8 years.”

If Bonnie and Clyde, heartless bank robbers that they are, care only about their own sentences, what would you expect them to do? Would they confess or remain silent? Figure 16-2 shows their choices. Each prisoner has two strategies: confess or remain silent. The sentence each prisoner gets depends on the strategy he or she chooses and the strategy chosen by his or her partner in crime.

Consider first Bonnie’s decision. She reasons as follows: “I don’t know what Clyde is going to do. If he remains silent, my best strategy is to confess, since then I’ll go free rather than spending a year in jail. If he confesses, my best strategy is

Figure 16-2

THE PRISONERS’ DILEMMA. In this game between two criminals suspected of committing a crime, the sentence that each receives depends both on his or her decision whether to confess or remain silent and on the decision made by the other.

		Bonnie's Decision	
		Confess	Remain Silent
Clyde's Decision	Confess	Bonnie gets 8 years Clyde gets 8 years	Bonnie gets 20 years Clyde goes free
	Remain Silent	Bonnie goes free Clyde gets 20 years	Bonnie gets 1 year Clyde gets 1 year

dominant strategy

a strategy that is best for a player in a game regardless of the strategies chosen by the other players

still to confess, since then I'll spend 8 years in jail rather than 20. So, regardless of what Clyde does, I am better off confessing."

In the language of game theory, a strategy is called a **dominant strategy** if it is the best strategy for a player to follow regardless of the strategies pursued by other players. In this case, confessing is a dominant strategy for Bonnie. She spends less time in jail if she confesses, regardless of whether Clyde confesses or remains silent.

Now consider Clyde's decision. He faces exactly the same choices as Bonnie, and he reasons in much the same way. Regardless of what Bonnie does, Clyde can reduce his time in jail by confessing. In other words, confessing is also a dominant strategy for Clyde.

In the end, both Bonnie and Clyde confess, and both spend 8 years in jail. Yet, from their standpoint, this is a terrible outcome. If they had *both* remained silent, both of them would have been better off, spending only 1 year in jail on the gun charge. By each pursuing his or her own interests, the two prisoners together reach an outcome that is worse for each of them.

To see how difficult it is to maintain cooperation, imagine that, before the police captured Bonnie and Clyde, the two criminals had made a pact not to confess. Clearly, this agreement would make them both better off *if* they both live up to it, because they would each spend only 1 year in jail. But would the two criminals in fact remain silent, simply because they had agreed to? Once they are being questioned separately, the logic of self-interest takes over and leads them to confess. Cooperation between the two prisoners is difficult to maintain, because cooperation is individually irrational.

OLIGOPOLIES AS A PRISONERS' DILEMMA

What does the prisoners' dilemma have to do with markets and imperfect competition? It turns out that the game oligopolists play in trying to reach the monopoly outcome is similar to the game that the two prisoners play in the prisoners' dilemma.

Consider an oligopoly with two members, called Iran and Iraq. Both countries sell crude oil. After prolonged negotiation, the countries agree to keep oil production low in order to keep the world price of oil high. After they agree on production levels, each country must decide whether to cooperate and live up to this agreement or to ignore it and produce at a higher level. Figure 16-3 shows how the profits of the two countries depend on the strategies they choose.

Suppose you are the president of Iraq. You might reason as follows: "I could keep production low as we agreed, or I could raise my production and sell more oil on world markets. If Iran lives up to the agreement and keeps its production low, then my country earns profit of \$60 billion with high production and \$50 billion with low production. In this case, Iraq is better off with high production. If Iran fails to live up to the agreement and produces at a high level, then my country earns \$40 billion with high production and \$30 billion with low production. Once again, Iraq is better off with high production. So, regardless of what Iran chooses to do, my country is better off reneging on our agreement and producing at a high level."

Producing at a high level is a dominant strategy for Iraq. Of course, Iran reasons in exactly the same way, and so both countries produce at a high level. The

Figure 16-3

AN OLIGOPOLY GAME. In this game between members of an oligopoly, the profit that each earns depends on both its production decision and the production decision of the other oligopolist.

		Iraq's Decision	
		High Production	Low Production
Iran's Decision	High Production	Iraq gets \$40 billion Iran gets \$40 billion	Iraq gets \$30 billion Iran gets \$60 billion
	Low Production	Iraq gets \$60 billion Iran gets \$30 billion	Iraq gets \$50 billion Iran gets \$50 billion

result is the inferior outcome (from Iran and Iraq's standpoint) with low profits for each country.

This example illustrates why oligopolies have trouble maintaining monopoly profits. The monopoly outcome is jointly rational for the oligopoly, but each oligopolist has an incentive to cheat. Just as self-interest drives the prisoners in the prisoners' dilemma to confess, self-interest makes it difficult for the oligopoly to maintain the cooperative outcome with low production, high prices, and monopoly profits.

OTHER EXAMPLES OF THE PRISONERS' DILEMMA

We have seen how the prisoners' dilemma can be used to understand the problem facing oligopolies. The same logic applies to many other situations as well. Here we consider three examples in which self-interest prevents cooperation and leads to an inferior outcome for the parties involved.

Arms Races An arms race is much like the prisoners' dilemma. To see this, consider the decisions of two countries—the United States and the Soviet Union—about whether to build new weapons or to disarm. Each country prefers to have more arms than the other because a larger arsenal gives it more influence in world affairs. But each country also prefers to live in a world safe from the other country's weapons.

Figure 16-4 shows the deadly game. If the Soviet Union chooses to arm, the United States is better off doing the same to prevent the loss of power. If the Soviet Union chooses to disarm, the United States is better off arming because doing so would make it more powerful. For each country, arming is a dominant strategy. Thus, each country chooses to continue the arms race, resulting in the inferior outcome in which both countries are at risk.

Figure 16-4

AN ARMS-RACE GAME. In this game between two countries, the safety and power of each country depends on both its decision whether to arm and the decision made by the other country.

		Decision of the United States (U.S.)	
		Arm	Disarm
Decision of the Soviet Union (USSR)	Arm	U.S. at risk USSR at risk	U.S. at risk and weak USSR safe and powerful
	Disarm	U.S. safe and powerful USSR at risk and weak	U.S. safe USSR safe

Throughout the era of the Cold War, the United States and the Soviet Union attempted to solve this problem through negotiation and agreements over arms control. The problems that the two countries faced were similar to those that oligopolists encounter in trying to maintain a cartel. Just as oligopolists argue over production levels, the United States and the Soviet Union argued over the amount of arms that each country would be allowed. And just as cartels have trouble enforcing production levels, the United States and the Soviet Union each feared that the other country would cheat on any agreement. In both arms races and oligopolies, the relentless logic of self-interest drives the participants toward a noncooperative outcome that is worse for each party.

Advertising When two firms advertise to attract the same customers, they face a problem similar to the prisoners' dilemma. For example, consider the decisions facing two cigarette companies, Marlboro and Camel. If neither company advertises, the two companies split the market. If both advertise, they again split the market, but profits are lower, since each company must bear the cost of advertising. Yet if one company advertises while the other does not, the one that advertises attracts customers from the other.

Figure 16-5 shows how the profits of the two companies depend on their actions. You can see that advertising is a dominant strategy for each firm. Thus, both firms choose to advertise, even though both firms would be better off if neither firm advertised.

A test of this theory of advertising occurred in 1971, when Congress passed a law banning cigarette advertisements on television. To the surprise of many observers, cigarette companies did not use their considerable political clout to oppose the law. When the law went into effect, cigarette advertising fell, and the profits of cigarette companies rose. The law did for the cigarette companies what they could not do on their own: It solved the prisoners' dilemma by enforcing the cooperative outcome with low advertising and high profit.

Figure 16-5

AN ADVERTISING GAME. In this game between firms selling similar products, the profit that each earns depends on both its own advertising decision and the advertising decision of the other firm.

		Marlboro's Decision	
		Advertise	Don't Advertise
Camel's Decision	Advertise	Marlboro gets \$3 billion profit Camel gets \$3 billion profit	Marlboro gets \$2 billion profit Camel gets \$5 billion profit
	Don't Advertise	Marlboro gets \$5 billion profit Camel gets \$2 billion profit	Marlboro gets \$4 billion profit Camel gets \$4 billion profit

Common Resources In Chapter 11 we saw that people tend to overuse common resources. One can view this problem as an example of the prisoners' dilemma.

Imagine that two oil companies—Exxon and Arco—own adjacent oil fields. Under the fields is a common pool of oil worth \$12 million. Drilling a well to recover the oil costs \$1 million. If each company drills one well, each will get half of the oil and earn a \$5 million profit (\$6 million in revenue minus \$1 million in costs).

Because the pool of oil is a common resource, the companies will not use it efficiently. Suppose that either company could drill a second well. If one company has two of the three wells, that company gets two-thirds of the oil, which yields a profit of \$6 million. Yet if each company drills a second well, the two companies again split the oil. In this case, each bears the cost of a second well, so profit is only \$4 million for each company.

Figure 16-6 shows the game. Drilling two wells is a dominant strategy for each company. Once again, the self-interest of the two players leads them to an inferior outcome.

THE PRISONERS' DILEMMA AND THE WELFARE OF SOCIETY

The prisoners' dilemma describes many of life's situations, and it shows that cooperation can be difficult to maintain, even when cooperation would make both players in the game better off. Clearly, this lack of cooperation is a problem for those involved in these situations. But is lack of cooperation a problem from the standpoint of society as a whole? The answer depends on the circumstances.

In some cases, the noncooperative equilibrium is bad for society as well as the players. In the arms-race game in Figure 16-4, both the United States and the

Figure 16-6

A COMMON-RESOURCES GAME. In this game between firms pumping oil from a common pool, the profit that each earns depends on both the number of wells it drills and the number of wells drilled by the other firm.

		Exxon's Decision	
		Drill Two Wells	Drill One Well
Arco's Decision	Drill Two Wells	Exxon gets \$4 million profit Arco gets \$4 million profit	Exxon gets \$3 million profit Arco gets \$6 million profit
	Drill One Well	Exxon gets \$6 million profit Arco gets \$3 million profit	Exxon gets \$5 million profit Arco gets \$5 million profit

Soviet Union end up at risk. In the common-resources game in Figure 16-6, the extra wells dug by Arco and Exxon are pure waste. In both cases, society would be better off if the two players could reach the cooperative outcome.

By contrast, in the case of oligopolists trying to maintain monopoly profits, lack of cooperation is desirable from the standpoint of society as a whole. The monopoly outcome is good for the oligopolists, but it is bad for the consumers of the product. As we first saw in Chapter 7, the competitive outcome is best for society because it maximizes total surplus. When oligopolists fail to cooperate, the quantity they produce is closer to this optimal level. Put differently, the invisible hand guides markets to allocate resources efficiently only when markets are competitive, and markets are competitive only when firms in the market fail to cooperate with one another.

Similarly, consider the case of the police questioning two suspects. Lack of cooperation between the suspects is desirable, for it allows the police to convict more criminals. The prisoners' dilemma is a dilemma for the prisoners, but it can be a boon to everyone else.

WHY PEOPLE SOMETIMES COOPERATE

The prisoners' dilemma shows that cooperation is difficult. But is it impossible? Not all prisoners, when questioned by the police, decide to turn in their partners in crime. Cartels sometimes do manage to maintain collusive arrangements, despite the incentive for individual members to defect. Very often, the reason that players can solve the prisoners' dilemma is that they play the game not once but many times.

To see why cooperation is easier to enforce in repeated games, let's return to our duopolists, Jack and Jill. Recall that Jack and Jill would like to maintain the monopoly outcome in which each produces 30 gallons, but self-interest drives

Figure 16-7

		Jack's Decision	
		Sell 40 Gallons	Sell 30 Gallons
Jill's Decision	Sell 40 Gallons	Jack gets \$1,600 profit Jill gets \$1,600 profit	Jack gets \$1,500 profit Jill gets \$2,000 profit
	Sell 30 Gallons	Jack gets \$2,000 profit Jill gets \$1,500 profit	Jack gets \$1,800 profit Jill gets \$1,800 profit

JACK AND JILL'S OLIGOPOLY GAME. In this game between Jack and Jill, the profit that each earns from selling water depends on both the quantity he or she chooses to sell and the quantity the other chooses to sell.

them to an equilibrium in which each produces 40 gallons. Figure 16-7 shows the game they play. Producing 40 gallons is a dominant strategy for each player in this game.

Imagine that Jack and Jill try to form a cartel. To maximize total profit, they would agree to the cooperative outcome in which each produces 30 gallons. Yet, if Jack and Jill are to play this game only once, neither has any incentive to live up to this agreement. Self-interest drives each of them to renege and produce 40 gallons.

Now suppose that Jack and Jill know that they will play the same game every week. When they make their initial agreement to keep production low, they can also specify what happens if one party reneges. They might agree, for instance, that once one of them reneges and produces 40 gallons, both of them will produce 40 gallons forever after. This penalty is easy to enforce, for if one party is producing at a high level, the other has every reason to do the same.

The threat of this penalty may be all that is needed to maintain cooperation. Each person knows that defecting would raise his or her profit from \$1,800 to \$2,000. But this benefit would last for only one week. Thereafter, profit would fall to \$1,600 and stay there. As long as the players care enough about future profits, they will choose to forgo the one-time gain from defection. Thus, in a game of repeated prisoners' dilemma, the two players may well be able to reach the cooperative outcome.

CASE STUDY THE PRISONERS' DILEMMA TOURNAMENT

Imagine that you are playing a game of prisoners' dilemma with a person being "questioned" in a separate room. Moreover, imagine that you are going to play not once but many times. Your score at the end of the game is the total number of years in jail. You would like to make this score as small as possible. What strategy would you play? Would you begin by confessing or remaining silent?

How would the other player's actions affect your subsequent decisions about confessing?

Repeated prisoners' dilemma is quite a complicated game. To encourage cooperation, players must penalize each other for not cooperating. Yet the strategy described earlier for Jack and Jill's water cartel—defect forever as soon as the other player defects—is not very forgiving. In a game repeated many times, a strategy that allows players to return to the cooperative outcome after a period of noncooperation may be preferable.

To see what strategies work best, political scientist Robert Axelrod held a tournament. People entered by sending computer programs designed to play repeated prisoners' dilemma. Each program then played the game against all the other programs. The "winner" was the program that received the fewest total years in jail.

The winner turned out to be a simple strategy called *tit-for-tat*. According to tit-for-tat, a player should start by cooperating and then do whatever the other player did last time. Thus, a tit-for-tat player cooperates until the other player defects; he then defects until the other player cooperates again. In other words, this strategy starts out friendly, penalizes unfriendly players, and forgives them if warranted. To Axelrod's surprise, this simple strategy did better than all the more complicated strategies that people had sent in.

The tit-for-tat strategy has a long history. It is essentially the biblical strategy of "an eye for an eye, a tooth for a tooth." The prisoners' dilemma tournament suggests that this may be a good rule of thumb for playing some of the games of life.

QUICK QUIZ: Tell the story of the prisoners' dilemma. Write down a table showing the prisoners' choices and explain what outcome is likely. ♦ What does the prisoners' dilemma teach us about oligopolies?

PUBLIC POLICY TOWARD OLIGOPOLIES



One of the *Ten Principles of Economics* in Chapter 1 is that governments can sometimes improve market outcomes. The application of this principle to oligopolistic markets is, as a general matter, straightforward. As we have seen, cooperation among oligopolists is undesirable from the standpoint of society as a whole, because it leads to production that is too low and prices that are too high. To move the allocation of resources closer to the social optimum, policymakers should try to induce firms in an oligopoly to compete rather than cooperate. Let's consider how policymakers do this and then examine the controversies that arise in this area of public policy.

RESTRAINT OF TRADE AND THE ANTITRUST LAWS

One way that policy discourages cooperation is through the common law. Normally, freedom of contract is an essential part of a market economy. Businesses and households use contracts to arrange mutually advantageous trades. In doing this,

they rely on the court system to enforce contracts. Yet, for many centuries, judges in England and the United States have deemed agreements among competitors to reduce quantities and raise prices to be contrary to the public good. They therefore refused to enforce such agreements.

The Sherman Antitrust Act of 1890 codified and reinforced this policy:

Every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations, is declared to be illegal. . . . Every person who shall monopolize, or attempt to monopolize, or combine or conspire with any person or persons to monopolize any part of the trade or commerce among the several States, or with foreign nations, shall be deemed guilty of a misdemeanor, and on conviction therefor, shall be punished by fine not exceeding fifty thousand dollars, or by imprisonment not exceeding one year, or by both said punishments, in the discretion of the court.

The Sherman Act elevated agreements among oligopolists from an unenforceable contract to a criminal conspiracy.

The Clayton Act of 1914 further strengthened the antitrust laws. According to this law, if a person could prove that he was damaged by an illegal arrangement to restrain trade, that person could sue and recover three times the damages he sustained. The purpose of this unusual rule of triple damages is to encourage private lawsuits against conspiring oligopolists.

Today, both the U.S. Justice Department and private parties have the authority to bring legal suits to enforce the antitrust laws. As we discussed in Chapter 15, these laws are used to prevent mergers that would lead to excessive market power in any single firm. In addition, these laws are used to prevent oligopolists from acting together in ways that would make their markets less competitive.

CASE STUDY AN ILLEGAL PHONE CALL

Firms in oligopolies have a strong incentive to collude in order to reduce production, raise price, and increase profit. The great eighteenth-century economist Adam Smith was well aware of this potential market failure. In *The Wealth of Nations* he wrote, “People of the same trade seldom meet together, but the conversation ends in a conspiracy against the public, or in some diversion to raise prices.”

To see a modern example of Smith’s observation, consider the following excerpt of a phone conversation between two airline executives in the early 1980s. The call was reported in *The New York Times* on February 24, 1983. Robert Crandall was president of American Airlines, and Howard Putnam was president of Braniff Airways.

CRANDALL: I think it’s dumb as hell . . . to sit here and pound the @#\$\$ out of each other and neither one of us making a #\$\$& dime.

PUTNAM: Do you have a suggestion for me?

CRANDALL: Yes, I have a suggestion for you. Raise your \$%*& fares 20 percent. I’ll raise mine the next morning.

PUTNAM: Robert, we . . .

CRANDALL: You’ll make more money, and I will, too.

IN THE NEWS*The Short Step from
Millionaire Executive
to Convicted Felon*

BUSINESS EXECUTIVES ARE SUPPOSED TO maximize their company's profits, but as the following article makes clear, they have to play within the rules established by the antitrust laws.

**Jury Convicts Ex-Executives
in ADM Case**

BY SCOTT KILMAN

CHICAGO—A federal jury found Michael D. Andreas and two other former Archer-

Daniels-Midland Co. executives guilty in a landmark price-fixing case.

The unanimous decision by the six-woman, six-man jury, reached here after a week of deliberations in the two-month trial, is a blow to the Andreas family, whose decades-long control of the Decatur, Ill., grain-processing giant has made it one of the Midwest's wealthiest and most politically influential families.

The verdicts also give the Justice Department its biggest convictions in a push against illegal global cartels. The department has 30 grand juries around the country considering international price-fixing cases, and more are expected. . . .

Mr. Andreas, who didn't take the stand in his defense, sat stone-faced as U.S. District Judge Blanche M. Manning read the verdict to the packed courtroom. Before the scandal, Mr. Andreas, 49 years old, was earning \$1.3 million

annually as the No. 2 executive at ADM and was being groomed to succeed his 80-year-old father, Dwayne Andreas. . . .

The most prominent American executive ever convicted for international price-fixing, the younger Mr. Andreas faces sentencing on Jan. 7. Prosecutors said they will seek the maximum sentence of three years in prison for violating the Sherman Antitrust Act. The jury determined that Mr. Andreas helped organize a cartel with four Asian companies to rig the \$650 million world-wide market for lysine, a fast-selling livestock-feed additive that hastens the growth of chickens and hogs. [Author's note: Andreas was eventually sentenced to spend two years in prison.]

SOURCE: *The Wall Street Journal*, September 18, 1998, p. A3.

PUTNAM: We can't talk about pricing!

CRANDALL: Oh @#\$%, Howard. We can talk about any &*\$@ thing we want to talk about.

Putnam was right: The Sherman Antitrust Act prohibits competing executives from even talking about fixing prices. When Howard Putnam gave a tape of this conversation to the Justice Department, the Justice Department filed suit against Robert Crandall.

Two years later, Crandall and the Justice Department reached a settlement in which Crandall agreed to various restrictions on his business activities, including his contacts with officials at other airlines. The Justice Department said that the terms of settlement would "protect competition in the airline industry, by preventing American and Crandall from any further attempts to monopolize passenger airline service on any route through discussions with competitors about the prices of airline services."

CONTROVERSIES OVER ANTITRUST POLICY

Over time, much controversy has centered on the question of what kinds of behavior the antitrust laws should prohibit. Most commentators agree that price-fixing agreements among competing firms should be illegal. Yet the antitrust laws

have been used to condemn some business practices whose effects are not obvious. Here we consider three examples.

Resale Price Maintenance One example of a controversial business practice is *resale price maintenance*, also called *fair trade*. Imagine that Superduper Electronics sells VCRs to retail stores for \$300. If Superduper requires the retailers to charge customers \$350, it is said to engage in resale price maintenance. Any retailer that charged less than \$350 would have violated its contract with Superduper.

At first, resale price maintenance might seem anticompetitive and, therefore, detrimental to society. Like an agreement among members of a cartel, it prevents the retailers from competing on price. For this reason, the courts have often viewed resale price maintenance as a violation of the antitrust laws.

Yet some economists defend resale price maintenance on two grounds. First, they deny that it is aimed at reducing competition. To the extent that Superduper Electronics has any market power, it can exert that power through the wholesale price, rather than through resale price maintenance. Moreover, Superduper has no incentive to discourage competition among its retailers. Indeed, because a cartel of retailers sells less than a group of competitive retailers, Superduper would be worse off if its retailers were a cartel.

Second, economists believe that resale price maintenance has a legitimate goal. Superduper may want its retailers to provide customers a pleasant showroom and a knowledgeable sales force. Yet, without resale price maintenance, some customers would take advantage of one store's service to learn about the VCR's special features and then buy the VCR at a discount retailer that does not provide this service. To some extent, good service is a public good among the retailers that sell Superduper VCRs. As we discussed in Chapter 11, when one person provides a public good, others are able to enjoy it without paying for it. In this case, discount retailers would free ride on the service provided by other retailers, leading to less service than is desirable. Resale price maintenance is one way for Superduper to solve this free-rider problem.

The example of resale price maintenance illustrates an important principle: *Business practices that appear to reduce competition may in fact have legitimate purposes.* This principle makes the application of the antitrust laws all the more difficult. The economists, lawyers, and judges in charge of enforcing these laws must determine what kinds of behavior public policy should prohibit as impeding competition and reducing economic well-being. Often that job is not easy.

Predatory Pricing Firms with market power normally use that power to raise prices above the competitive level. But should policymakers ever be concerned that firms with market power might charge prices that are too low? This question is at the heart of a second debate over antitrust policy.

Imagine that a large airline, call it Coyote Air, has a monopoly on some route. Then Roadrunner Express enters and takes 20 percent of the market, leaving Coyote with 80 percent. In response to this competition, Coyote starts slashing its fares. Some antitrust analysts argue that Coyote's move could be anticompetitive: The price cuts may be intended to drive Roadrunner out of the market so Coyote can recapture its monopoly and raise prices again. Such behavior is called *predatory pricing*.

Although predatory pricing is a common claim in antitrust suits, some economists are skeptical of this argument and believe that predatory pricing is rarely,

and perhaps never, a profitable business strategy. Why? For a price war to drive out a rival, prices have to be driven below cost. Yet if Coyote starts selling cheap tickets at a loss, it had better be ready to fly more planes, because low fares will attract more customers. Roadrunner, meanwhile, can respond to Coyote's predatory move by cutting back on flights. As a result, Coyote ends up bearing more than 80 percent of the losses, putting Roadrunner in a good position to survive the price war. As in the old Roadrunner–Coyote cartoons, the predator suffers more than the prey.

Economists continue to debate whether predatory pricing should be a concern for antitrust policymakers. Various questions remain unresolved. Is predatory pricing ever a profitable business strategy? If so, when? Are the courts capable of telling which price cuts are competitive and thus good for consumers and which are predatory? There are no easy answers.

Tying A third example of a controversial business practice is *tying*. Suppose that Makemoney Movies produces two new films—*Star Wars* and *Hamlet*. If Makemoney offers theaters the two films together at a single price, rather than separately, the studio is said to be tying its two products.

When the practice of tying movies was challenged in the courts, the U.S. Supreme Court banned the practice. The Court reasoned as follows: Imagine that *Star Wars* is a blockbuster, whereas *Hamlet* is an unprofitable art film. Then the studio could use the high demand for *Star Wars* to force theaters to buy *Hamlet*. It seemed that the studio could use tying as a mechanism for expanding its market power.

Many economists, however, are skeptical of this argument. Imagine that theaters are willing to pay \$20,000 for *Star Wars* and nothing for *Hamlet*. Then the most that a theater would pay for the two movies together is \$20,000—the same as it would pay for *Star Wars* by itself. Forcing the theater to accept a worthless movie as part of the deal does not increase the theater's willingness to pay. Makemoney cannot increase its market power simply by bundling the two movies together.

Why, then, does tying exist? One possibility is that it is a form of price discrimination. Suppose there are two theaters. City Theater is willing to pay \$15,000 for *Star Wars* and \$5,000 for *Hamlet*. Country Theater is just the opposite: It is willing to pay \$5,000 for *Star Wars* and \$15,000 for *Hamlet*. If Makemoney charges separate prices for the two films, its best strategy is to charge \$15,000 for each film, and each theater chooses to show only one film. Yet if Makemoney offers the two movies as a bundle, it can charge each theater \$20,000 for the movies. Thus, if different theaters value the films differently, tying may allow the studio to increase profit by charging a combined price closer to the buyers' total willingness to pay.

Tying remains a controversial business practice. The Supreme Court's argument that tying allows a firm to extend its market power to other goods is not well founded, at least in its simplest form. Yet economists have proposed more elaborate theories for how tying can impede competition. Given our current economic knowledge, it is unclear whether tying has adverse effects for society as a whole.

CASE STUDY THE MICROSOFT CASE

The most important and controversial antitrust case in recent years has been the U.S. government's suit against the Microsoft Corporation, filed in 1998. Certainly, the case did not lack drama. It pitted one of the world's richest men (Bill Gates) against one of the world's most powerful regulatory agencies (the U.S.

Justice Department). Testifying for the government was a prominent economist (MIT professor Franklin Fisher). Testifying for Microsoft was an equally prominent economist (MIT professor Richard Schmalensee). At stake was the future of one of the world's most valuable companies (Microsoft) in one of the economy's fastest growing industries (computer software).

A central issue in the Microsoft case involved tying—in particular, whether Microsoft should be allowed to integrate its Internet browser into its Windows operating system. The government claimed that Microsoft was bundling these two products together to expand the market power it had in the market for computer operating systems into an unrelated market (for Internet browsers). Allowing Microsoft to incorporate such products into its operating system, the government argued, would deter new software companies such as Netscape from entering the market and offering new products.

Microsoft responded by pointing out that putting new features into old products is a natural part of technological progress. Cars today include stereos and air-conditioners, which were once sold separately, and cameras come with built-in flashes. The same is true with operating systems. Over time, Microsoft has added many features to Windows that were previously stand-alone products. This has made computers more reliable and easier to use because consumers can be confident that the pieces work together. The integration of Internet technology, Microsoft argued, was the natural next step.

One point of disagreement concerned the extent of Microsoft's market power. Noting that more than 80 percent of new personal computers used a Microsoft operating system, the government argued that the company had substantial monopoly power, which it was trying to expand. Microsoft replied that the software market is always changing and that Microsoft's Windows was constantly being challenged by competitors, such as the Apple Mac and Linux operating systems. It also argued that the low price it charged for Windows—about \$50, or only 3 percent of the price of a typical computer—was evidence that its market power was severely limited.

As this book was going to press, the final outcome of the Microsoft case was yet to be resolved. In November 1999 the trial judge issued a ruling in which he found that Microsoft had great monopoly power and that it had illegally abused that power. But many questions were still unanswered. Would the trial court's decision hold up on appeal? If so, what remedy would the government seek? Would it try to regulate future design changes in the Windows operating system? Would it try to break up Microsoft into a group of smaller, more competitive companies? The answers to these questions will shape the software industry for years to come.



"Me? A MONOPOLIST? NOW JUST WAIT A MINUTE . . ."

QUICK QUIZ: What kind of agreement is illegal for businesses to make?

◆ Why are the antitrust laws controversial?

CONCLUSION

Oligopolies would like to act like monopolies, but self-interest drives them closer to competition. Thus, oligopolies can end up looking either more like monopolies or more like competitive markets, depending on the number of firms in the

oligopoly and how cooperative the firms are. The story of the prisoners' dilemma shows why oligopolies can fail to maintain cooperation, even when cooperation is in their best interest.

Policymakers regulate the behavior of oligopolists through the antitrust laws. The proper scope of these laws is the subject of ongoing controversy. Although price fixing among competing firms clearly reduces economic welfare and should be illegal, some business practices that appear to reduce competition may have legitimate if subtle purposes. As a result, policymakers need to be careful when they use the substantial powers of the antitrust laws to place limits on firm behavior.

Summary

- ◆ Oligopolists maximize their total profits by forming a cartel and acting like a monopolist. Yet, if oligopolists make decisions about production levels individually, the result is a greater quantity and a lower price than under the monopoly outcome. The larger the number of firms in the oligopoly, the closer the quantity and price will be to the levels that would prevail under competition.
- ◆ The prisoners' dilemma shows that self-interest can prevent people from maintaining cooperation, even when cooperation is in their mutual interest. The logic of the prisoners' dilemma applies in many situations, including arms races, advertising, common-resource problems, and oligopolies.
- ◆ Policymakers use the antitrust laws to prevent oligopolies from engaging in behavior that reduces competition. The application of these laws can be controversial, because some behavior that may seem to reduce competition may in fact have legitimate business purposes.

Key Concepts

oligopoly, p. 350
monopolistic competition, p. 350
collusion, p. 353

cartel, p. 353
Nash equilibrium, p. 355
game theory, p. 358

prisoners' dilemma, p. 359
dominant strategy, p. 360

Questions for Review

1. If a group of sellers could form a cartel, what quantity and price would they try to set?
2. Compare the quantity and price of an oligopoly to those of a monopoly.
3. Compare the quantity and price of an oligopoly to those of a competitive market.
4. How does the number of firms in an oligopoly affect the outcome in its market?
5. What is the prisoners' dilemma, and what does it have to do with oligopoly?
6. Give two examples other than oligopoly to show how the prisoners' dilemma helps to explain behavior.
7. What kinds of behavior do the antitrust laws prohibit?
8. What is resale price maintenance, and why is it controversial?

Problems and Applications

1. *The New York Times* (Nov. 30, 1993) reported that “the inability of OPEC to agree last week to cut production has sent the oil market into turmoil . . . [leading to] the lowest price for domestic crude oil since June 1990.”
 - a. Why were the members of OPEC trying to agree to cut production?
 - b. Why do you suppose OPEC was unable to agree on cutting production? Why did the oil market go into “turmoil” as a result?
 - c. The newspaper also noted OPEC’s view “that producing nations outside the organization, like Norway and Britain, should do their share and cut production.” What does the phrase “do their share” suggest about OPEC’s desired relationship with Norway and Britain?
2. A large share of the world supply of diamonds comes from Russia and South Africa. Suppose that the marginal cost of mining diamonds is constant at \$1,000 per diamond, and the demand for diamonds is described by the following schedule:

PRICE	QUANTITY
\$8,000	5,000
7,000	6,000
6,000	7,000
5,000	8,000
4,000	9,000
3,000	10,000
2,000	11,000
1,000	12,000

- a. If there were many suppliers of diamonds, what would be the price and quantity?
 - b. If there were only one supplier of diamonds, what would be the price and quantity?
 - c. If Russia and South Africa formed a cartel, what would be the price and quantity? If the countries split the market evenly, what would be South Africa’s production and profit? What would happen to South Africa’s profit if it increased its production by 1,000 while Russia stuck to the cartel agreement?
 - d. Use your answer to part (c) to explain why cartel agreements are often not successful.
3. This chapter discusses companies that are oligopolists in the market for the goods they sell. Many of the same

ideas apply to companies that are oligopolists in the market for the inputs they buy.

- a. If sellers who are oligopolists try to increase the price of goods they sell, what is the goal of buyers who are oligopolists?
 - b. Major league baseball team owners have an oligopoly in the market for baseball players. What is the owners’ goal regarding players’ salaries? Why is this goal difficult to achieve?
 - c. Baseball players went on strike in 1994 because they would not accept the salary cap that the owners wanted to impose. If the owners were already colluding over salaries, why did the owners feel the need for a salary cap?
4. Describe several activities in your life in which game theory could be useful. What is the common link among these activities?
 5. Consider trade relations between the United States and Mexico. Assume that the leaders of the two countries believe the payoffs to alternative trade policies are as follows:

		United States’ Decision	
		Low Tariffs	High Tariffs
Mexico’s Decision	Low Tariffs	U.S. gains \$25 billion Mexico gains \$25 billion	U.S. gains \$30 billion Mexico gains \$10 billion
	High Tariffs	U.S. gains \$10 billion Mexico gains \$30 billion	U.S. gains \$20 billion Mexico gains \$20 billion

- a. What is the dominant strategy for the United States? For Mexico? Explain.
- b. Define *Nash equilibrium*. What is the Nash equilibrium for trade policy?
- c. In 1993 the U.S. Congress ratified the North American Free Trade Agreement (NAFTA), in which the United States and Mexico agreed to reduce trade barriers simultaneously. Do the perceived payoffs as shown here justify this approach to trade policy?
- d. Based on your understanding of the gains from trade (discussed in Chapters 3 and 9), do you think that these payoffs actually reflect a nation’s welfare under the four possible outcomes?

6. Suppose that you and a classmate are assigned a project on which you will receive one combined grade. You each want to receive a good grade, but you also want to do as little work as possible. The decision box and payoffs are as follows:

		Your Decision	
		Work	Shirk
Classmate's Decision	Work	You get A grade, no fun Classmate gets A grade, no fun	You get B grade, fun Classmate gets B grade, no fun
	Shirk	You get B grade, no fun Classmate gets B grade, fun	You get D grade, fun Classmate gets D grade, fun

Assume that having fun is your normal state, but having no fun is as unpleasant as receiving a grade that is two letters lower.

- Write out the decision box that combines the letter grade and the amount of fun you have into a single payoff for each outcome.
 - If neither you nor your classmate knows how much work the other person is doing, what is the likely outcome? Does it matter whether you are likely to work with this person again? Explain your answer.
7. The chapter states that the ban on cigarette advertising on television in 1971 increased the profits of cigarette companies. Could the ban still be good public policy? Explain your answer.
8. A case study in the chapter describes a phone conversation between the presidents of American Airlines and Braniff Airways. Let's analyze the game between the two companies. Suppose that each company can charge either a high price for tickets or a low price. If one company charges \$100, it earns low profits if the other company charges \$100 also, and high profits if the other company charges \$200. On the other hand, if the company charges \$200, it earns very low profits if the other company charges \$100, and medium profits if the other company charges \$200 also.
- Draw the decision box for this game.
 - What is the Nash equilibrium in this game? Explain.
 - Is there an outcome that would be better than the Nash equilibrium for both airlines? How could it be achieved? Who would lose if it were achieved?

9. Farmer Jones and Farmer Smith graze their cattle on the same field. If there are 20 cows grazing in the field, each cow produces \$4,000 of milk over its lifetime. If there are more cows in the field, then each cow can eat less grass, and its milk production falls. With 30 cows on the field, each produces \$3,000 of milk; with 40 cows, each produces \$2,000 of milk. Cows cost \$1,000 apiece.

- Assume that Farmer Jones and Farmer Smith can each purchase either 10 or 20 cows, but that neither knows how many the other is buying when she makes her purchase. Calculate the payoffs of each outcome.
 - What is the likely outcome of this game? What would be the best outcome? Explain.
 - There used to be more common fields than there are today. Why? (For more discussion of this topic, reread Chapter 11.)
10. Little Kona is a small coffee company that is considering entering a market dominated by Big Brew. Each company's profit depends on whether Little Kona enters and whether Big Brew sets a high price or a low price:

		Big Brew	
		High Price	Low Price
Little Kona	Enter	Brew makes \$3 million Kona makes \$2 million	Brew makes \$1 million Kona loses \$1 million
	Don't Enter	Brew makes \$7 million Kona makes zero	Brew makes \$2 million Kona makes zero

Big Brew threatens Little Kona by saying, "If you enter, we're going to set a low price, so you had better stay out." Do you think Little Kona should believe the threat? Why or why not? What do you think Little Kona should do?

11. Jeff and Steve are playing tennis. Every point comes down to whether Steve guesses correctly whether Jeff will hit the ball to Steve's left or right. The outcomes are:

		Steve Guesses	
		Left	Right
Jeff Hits	Left	Jeff loses point Steve wins point	Jeff wins point Steve loses point
	Right	Jeff wins point Steve loses point	Jeff loses point Steve wins point

Does either player have a dominant strategy? If Jeff chooses a particular strategy (Left or Right) and sticks with it, what will Steve do? So, can you think of a better strategy for Jeff to follow?

17



MONOPOLISTIC COMPETITION

You walk into a bookstore to buy a book to read during your next vacation. On the store's shelves you find a John Grisham mystery, a Stephen King thriller, a Danielle Steel romance, a Frank McCourt memoir, and many other choices. When you pick out a book and buy it, what kind of market are you participating in?

On the one hand, the market for books seems competitive. As you look over the shelves at your bookstore, you find many authors and many publishers vying for your attention. A buyer in this market has thousands of competing products from which to choose. And because anyone can enter the industry by writing and publishing a book, the book business is not very profitable. For every highly paid novelist, there are hundreds of struggling ones.

On the other hand, the market for books seems monopolistic. Because each book is unique, publishers have some latitude in choosing what price to charge. The sellers in this market are price makers rather than price takers. And, indeed, the price of books greatly exceeds marginal cost. The price of a typical hardcover

IN THIS CHAPTER YOU WILL . . .

Analyze competition among firms that sell differentiated products

Compare the outcome under monopolistic competition and under perfect competition

Consider the desirability of outcomes in monopolistically competitive markets

Examine the debate over the effects of advertising

Examine the debate over the role of brand names

monopolistic competition

a market structure in which many firms sell products that are similar but not identical

novel, for instance, is about \$25, whereas the cost of printing one additional copy of the novel is less than \$5.

In this chapter we examine markets that have some features of competition and some features of monopoly. This market structure is called **monopolistic competition**. Monopolistic competition describes a market with the following attributes:

- ◆ *Many sellers:* There are many firms competing for the same group of customers.
- ◆ *Product differentiation:* Each firm produces a product that is at least slightly different from those of other firms. Thus, rather than being a price taker, each firm faces a downward-sloping demand curve.
- ◆ *Free entry:* Firms can enter (or exit) the market without restriction. Thus, the number of firms in the market adjusts until economic profits are driven to zero.

A moment's thought reveals a long list of markets with these attributes: books, CDs, movies, computer games, restaurants, piano lessons, cookies, furniture, and so on.

Monopolistic competition, like oligopoly, is a market structure that lies between the extreme cases of competition and monopoly. But oligopoly and monopolistic competition are quite different. Oligopoly departs from the perfectly competitive ideal of Chapter 14 because there are only a few sellers in the market. The small number of sellers makes rigorous competition less likely, and it makes strategic interactions among them vitally important. By contrast, under monopolistic competition, there are many sellers, each of which is small compared to the market. A monopolistically competitive market departs from the perfectly competitive ideal because each of the sellers offers a somewhat different product.

COMPETITION WITH DIFFERENTIATED PRODUCTS

To understand monopolistically competitive markets, we first consider the decisions facing an individual firm. We then examine what happens in the long run as firms enter and exit the industry. Next, we compare the equilibrium under monopolistic competition to the equilibrium under perfect competition that we examined in Chapter 14. Finally, we consider whether the outcome in a monopolistically competitive market is desirable from the standpoint of society as a whole.

THE MONOPOLISTICALLY COMPETITIVE FIRM IN THE SHORT RUN

Each firm in a monopolistically competitive market is, in many ways, like a monopoly. Because its product is different from those offered by other firms, it faces a

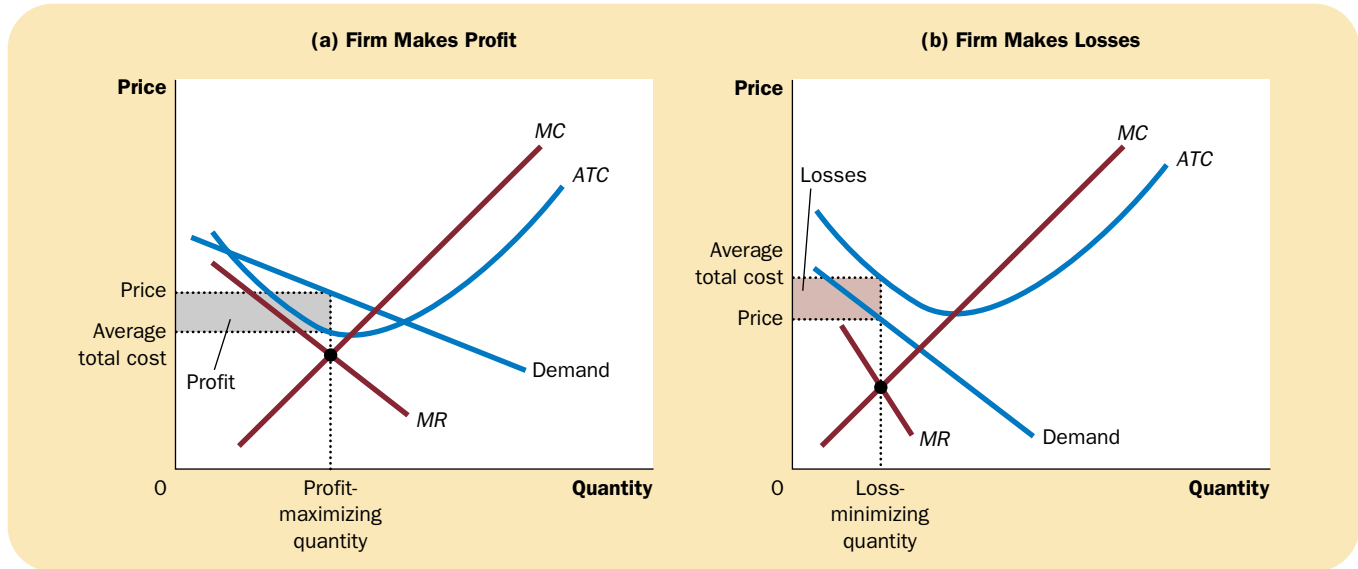


Figure 17-1

MONOPOLISTIC COMPETITORS IN THE SHORT RUN. Monopolistic competitors, like monopolists, maximize profit by producing the quantity at which marginal revenue equals marginal cost. The firm in panel (a) makes a profit because, at this quantity, price is above average total cost. The firm in panel (b) makes losses because, at this quantity, price is less than average total cost.

downward-sloping demand curve. (By contrast, a perfectly competitive firm faces a horizontal demand curve at the market price.) Thus, the monopolistically competitive firm follows a monopolist's rule for profit maximization: It chooses the quantity at which marginal revenue equals marginal cost and then uses its demand curve to find the price consistent with that quantity.

Figure 17-1 shows the cost, demand, and marginal-revenue curves for two typical firms, each in a different monopolistically competitive industry. In both panels of this figure, the profit-maximizing quantity is found at the intersection of the marginal-revenue and marginal-cost curves. The two panels in this figure show different outcomes for the firm's profit. In panel (a), price exceeds average total cost, so the firm makes a profit. In panel (b), price is below average total cost. In this case, the firm is unable to make a positive profit, so the best the firm can do is to minimize its losses.

All this should seem familiar. A monopolistically competitive firm chooses its quantity and price just as a monopoly does. In the short run, these two types of market structure are similar.

THE LONG-RUN EQUILIBRIUM

The situations depicted in Figure 17-1 do not last long. When firms are making profits, as in panel (a), new firms have an incentive to enter the market. This



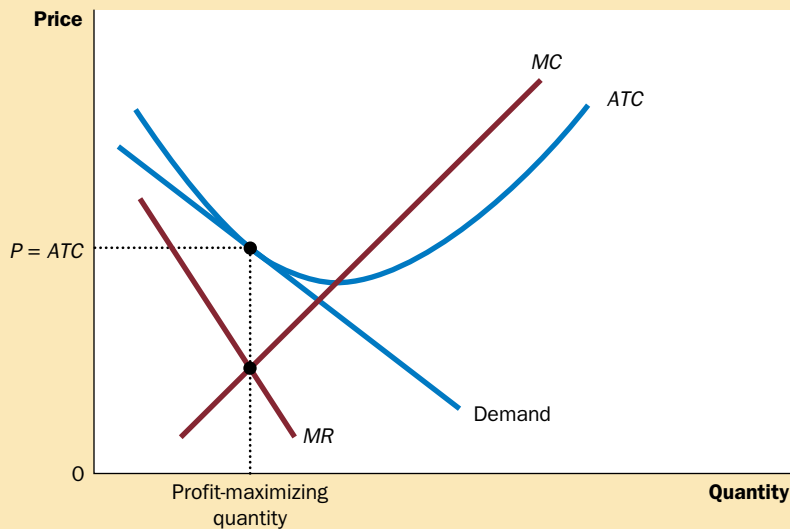
"GIVEN THE DOWNWARD SLOPE OF OUR DEMAND CURVE, AND THE EASE WITH WHICH OTHER FIRMS CAN ENTER THE INDUSTRY, WE CAN STRENGTHEN OUR PROFIT POSITION ONLY BY EQUATING MARGINAL COST AND MARGINAL REVENUE. ORDER MORE JELLY BEANS."

entry increases the number of products from which customers can choose and, therefore, reduces the demand faced by each firm already in the market. In other words, profit encourages entry, and entry shifts the demand curves faced by the incumbent firms to the left. As the demand for incumbent firms' products falls, these firms experience declining profit.

Conversely, when firms are making losses, as in panel (b), firms in the market have an incentive to exit. As firms exit, customers have fewer products from which to choose. This decrease in the number of firms expands the demand faced by those firms that stay in the market. In other words, losses encourage exit, and exit shifts the demand curves of the remaining firms to the right. As the demand for the remaining firms' products rises, these firms experience rising profit (that is, declining losses).

This process of entry and exit continues until the firms in the market are making exactly zero economic profit. Figure 17-2 depicts the long-run equilibrium. Once the market reaches this equilibrium, new firms have no incentive to enter, and existing firms have no incentive to exit.

Notice that the demand curve in this figure just barely touches the average-total-cost curve. Mathematically, we say the two curves are *tangent* to each other. These two curves must be tangent once entry and exit have driven profit to zero. Because profit per unit sold is the difference between price (found on the demand curve) and average total cost, the maximum profit is zero only if these two curves touch each other without crossing.

**Figure 17-2**

A MONOPOLISTIC COMPETITOR IN THE LONG RUN. In a monopolistically competitive market, if firms are making profit, new firms enter, and the demand curves for the incumbent firms shift to the left. Similarly, if firms are making losses, old firms exit, and the demand curves of the remaining firms shift to the right. Because of these shifts in demand, a monopolistically competitive firm eventually finds itself in the long-run equilibrium shown here. In this long-run equilibrium, price equals average total cost, and the firm earns zero profit.

To sum up, two characteristics describe the long-run equilibrium in a monopolistically competitive market:

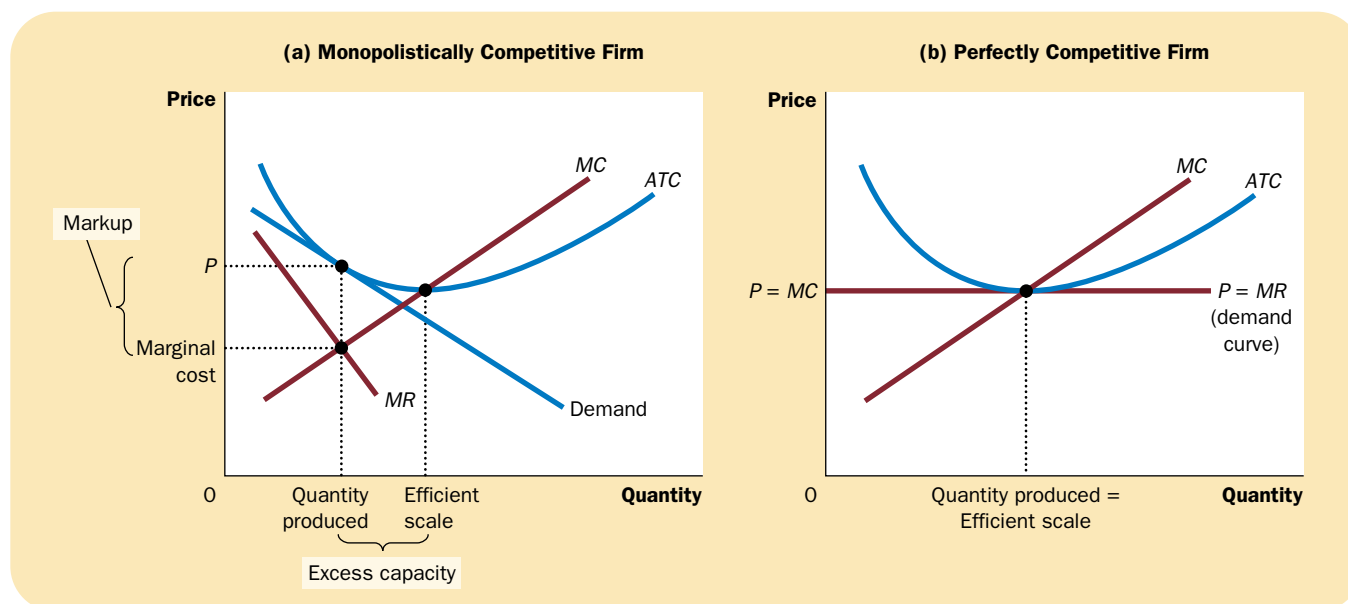
- ◆ As in a monopoly market, price exceeds marginal cost. This conclusion arises because profit maximization requires marginal revenue to equal marginal cost and because the downward sloping demand curve makes marginal revenue less than the price.
- ◆ As in a competitive market, price equals average total cost. This conclusion arises because free entry and exit drive economic profit to zero.

The second characteristic shows how monopolistic competition differs from monopoly. Because a monopoly is the sole seller of a product without close substitutes, it can earn positive economic profit, even in the long run. By contrast, because there is free entry into a monopolistically competitive market, the economic profit of a firm in this type of market is driven to zero.

MONOPOLISTIC VERSUS PERFECT COMPETITION

Figure 17-3 compares the long-run equilibrium under monopolistic competition to the long-run equilibrium under perfect competition. (Chapter 14 discussed the equilibrium with perfect competition.) There are two noteworthy differences between monopolistic and perfect competition: excess capacity and the markup.

Excess Capacity As we have just seen, entry and exit drive each firm in a monopolistically competitive market to a point of tangency between its demand

**Figure 17-3**

MONOPOLISTIC VERSUS PERFECT COMPETITION. Panel (a) shows the long-run equilibrium in a monopolistically competitive market, and panel (b) shows the long-run equilibrium in a perfectly competitive market. Two differences are notable. (1) The perfectly competitive firm produces at the efficient scale, where average total cost is minimized. By contrast, the monopolistically competitive firm produces at less than the efficient scale. (2) Price equals marginal cost under perfect competition, but price is above marginal cost under monopolistic competition.

and average-total-cost curves. Panel (a) of Figure 17-3 shows that the quantity of output at this point is smaller than the quantity that minimizes average total cost. Thus, under monopolistic competition, firms produce on the downward-sloping portion of their average-total-cost curves. In this way, monopolistic competition contrasts starkly with perfect competition. As panel (b) of Figure 17-3 shows, free entry in competitive markets drives firms to produce at the minimum of average total cost.

The quantity that minimizes average total cost is called the *efficient scale* of the firm. In the long run, perfectly competitive firms produce at the efficient scale, whereas monopolistically competitive firms produce below this level. Firms are said to have *excess capacity* under monopolistic competition. In other words, a monopolistically competitive firm, unlike a perfectly competitive firm, could increase the quantity it produces and lower the average total cost of production.

Markup over Marginal Cost A second difference between perfect competition and monopolistic competition is the relationship between price and marginal cost. For a competitive firm, such as that shown in panel (b) of Figure 17-3, price equals marginal cost. For a monopolistically competitive firm, such as that shown in panel (a), price exceeds marginal cost, because the firm always has some market power.

How is this markup over marginal cost consistent with free entry and zero profit? The zero-profit condition ensures only that price equals average total cost. It does *not* ensure that price equals marginal cost. Indeed, in the long-run equilibrium, monopolistically competitive firms operate on the declining portion of their average-total-cost curves, so marginal cost is below average total cost. Thus, for price to equal average total cost, price must be above marginal cost.

In this relationship between price and marginal cost, we see a key behavioral difference between perfect competitors and monopolistic competitors. Imagine that you were to ask a firm the following question: “Would you like to see another customer come through your door ready to buy from you at your current price?” A perfectly competitive firm would answer that it didn’t care. Because price exactly equals marginal cost, the profit from an extra unit sold is zero. By contrast, a monopolistically competitive firm is always eager to get another customer. Because its price exceeds marginal cost, an extra unit sold at the posted price means more profit. According to an old quip, monopolistically competitive markets are those in which sellers send Christmas cards to the buyers.

MONOPOLISTIC COMPETITION AND THE WELFARE OF SOCIETY

Is the outcome in a monopolistically competitive market desirable from the standpoint of society as a whole? Can policymakers improve on the market outcome? There are no simple answers to these questions.

One source of inefficiency is the markup of price over marginal cost. Because of the markup, some consumers who value the good at more than the marginal cost of production (but less than the price) will be deterred from buying it. Thus, a monopolistically competitive market has the normal deadweight loss of monopoly pricing. We first saw this type of inefficiency when we discussed monopoly in Chapter 15.

Although this outcome is clearly undesirable compared to the first-best outcome of price equal to marginal cost, there is no easy way for policymakers to fix the problem. To enforce marginal-cost pricing, policymakers would need to regulate all firms that produce differentiated products. Because such products are so common in the economy, the administrative burden of such regulation would be overwhelming.

Moreover, regulating monopolistic competitors would entail all the problems of regulating natural monopolies. In particular, because monopolistic competitors are making zero profits already, requiring them to lower their prices to equal marginal cost would cause them to make losses. To keep these firms in business, the government would need to help them cover these losses. Rather than raising taxes to pay for these subsidies, policymakers may decide it is better to live with the inefficiency of monopolistic pricing.

Another way in which monopolistic competition may be socially inefficient is that the number of firms in the market may not be the “ideal” one. That is, there may be too much or too little entry. One way to think about this problem is in terms of the externalities associated with entry. Whenever a new firm considers entering the market with a new product, it considers only the profit it would make. Yet its entry would also have two external effects:

- ◆ *The product-variety externality:* Because consumers get some consumer surplus from the introduction of a new product, entry of a new firm conveys a positive externality on consumers.
- ◆ *The business-stealing externality:* Because other firms lose customers and profits from the entry of a new competitor, entry of a new firm imposes a negative externality on existing firms.

Thus, in a monopolistically competitive market, there are both positive and negative externalities associated with the entry of new firms. Depending on which externality is larger, a monopolistically competitive market could have either too few or too many products.

Both of these externalities are closely related to the conditions for monopolistic competition. The product-variety externality arises because a new firm would offer a product different from those of the existing firms. The business-stealing externality arises because firms post a price above marginal cost and, therefore, are always eager to sell additional units. Conversely, because perfectly competitive firms produce identical goods and charge a price equal to marginal cost, neither of these externalities exists under perfect competition.

In the end, we can conclude only that monopolistically competitive markets do not have all the desirable welfare properties of perfectly competitive markets. That is, the invisible hand does not ensure that total surplus is maximized under monopolistic competition. Yet because the inefficiencies are subtle, hard to measure, and hard to fix, there is no easy way for public policy to improve the market outcome.

QUICK QUIZ: List the three key attributes of monopolistic competition.

- ◆ Draw and explain a diagram to show the long-run equilibrium in a monopolistically competitive market. How does this equilibrium differ from that in a perfectly competitive market?

FYI

Is Excess Capacity a Social Problem?



monopolistic competitors was a source of inefficiency.

Today economists understand that the excess capacity of monopolistic competitors is not directly relevant for

As we have seen, monopolistically competitive firms produce a quantity of output below the level that minimizes average total cost. By contrast, firms in perfectly competitive markets are driven to produce at the quantity that minimizes average total cost. This comparison between perfect and monopolistic competition has led some economists in the past to argue that the excess capacity of

evaluating economic welfare. There is no reason that society should want all firms to produce at the minimum of average total cost. For example, consider a publishing firm. Producing a novel might take a fixed cost of \$50,000 (the author's time) and variable costs of \$5 per book (the cost of printing). In this case, the average total cost of a book declines as the number of books increases because the fixed cost gets spread over more and more units. The average total cost is minimized by printing an infinite number of books. But in no sense is infinity the right number of books for society to produce.

In short, monopolistic competitors do have excess capacity, but this fact tells us little about the desirability of the market outcome.

ADVERTISING

It is nearly impossible to go through a typical day in a modern economy without being bombarded with advertising. Whether you are reading a newspaper, watching television, or driving down the highway, some firm will try to convince you to buy its product. Such behavior is a natural feature of monopolistic competition. When firms sell differentiated products and charge prices above marginal cost, each firm has an incentive to advertise in order to attract more buyers to its particular product.

The amount of advertising varies substantially across products. Firms that sell highly differentiated consumer goods, such as over-the-counter drugs, perfumes, soft drinks, razor blades, breakfast cereals, and dog food, typically spend between 10 and 20 percent of revenue for advertising. Firms that sell industrial products, such as drill presses and communications satellites, typically spend very little on advertising. And firms that sell homogeneous products, such as wheat, peanuts, or crude oil, spend nothing at all. For the economy as a whole, spending on advertising comprises about 2 percent of total firm revenue, or more than \$100 billion.

Advertising takes many forms. About one-half of advertising spending is for space in newspapers and magazines, and about one-third is for commercials on television and radio. The rest is spent on various other ways of reaching customers, such as direct mail, billboards, and the Goodyear blimp.

THE DEBATE OVER ADVERTISING

Is society wasting the resources it devotes to advertising? Or does advertising serve a valuable purpose? Assessing the social value of advertising is difficult and often generates heated argument among economists. Let's consider both sides of the debate.

The Critique of Advertising Critics of advertising argue that firms advertise in order to manipulate people's tastes. Much advertising is psychological rather than informational. Consider, for example, the typical television commercial for some brand of soft drink. The commercial most likely does not tell the viewer about the product's price or quality. Instead, it might show a group of happy people at a party on a beach on a beautiful sunny day. In their hands are cans of the soft drink. The goal of the commercial is to convey a subconscious (if not subtle) message: "You too can have many friends and be happy, if only you drink our product." Critics of advertising argue that such a commercial creates a desire that otherwise might not exist.

Critics also argue that advertising impedes competition. Advertising often tries to convince consumers that products are more different than they truly are. By increasing the perception of product differentiation and fostering brand loyalty, advertising makes buyers less concerned with price differences among similar goods. With a less elastic demand curve, each firm charges a larger markup over marginal cost.

The Defense of Advertising Defenders of advertising argue that firms use advertising to provide information to customers. Advertising conveys the

prices of the goods being offered for sale, the existence of new products, and the locations of retail outlets. This information allows customers to make better choices about what to buy and, thus, enhances the ability of markets to allocate resources efficiently.

Defenders also argue that advertising fosters competition. Because advertising allows customers to be more fully informed about all the firms in the market, customers can more easily take advantage of price differences. Thus, each firm has less market power. In addition, advertising allows new firms to enter more easily, because it gives entrants a means to attract customers from existing firms.

Over time, policymakers have come to accept the view that advertising can make markets more competitive. One important example is the regulation of certain professions, such as lawyers, doctors, and pharmacists. In the past, these groups succeeded in getting state governments to prohibit advertising in their fields on the grounds that advertising was “unprofessional.” In recent years, however, the courts have concluded that the primary effect of these restrictions on advertising was to curtail competition. They have, therefore, overturned many of the laws that prohibit advertising by members of these professions.

CASE STUDY ADVERTISING AND THE PRICE OF EYEGLASSES

What effect does advertising have on the price of a good? On the one hand, advertising might make consumers view products as being more different than they otherwise would. If so, it would make markets less competitive and firms’ demand curves less elastic, and this would lead firms to charge higher prices. On the other hand, advertising might make it easier for consumers to find the firms offering the best prices. In this case, it would make markets more competitive and firms’ demand curves more elastic, and this would lead to lower prices.

In an article published in the *Journal of Law and Economics* in 1972, economist Lee Benham tested these two views of advertising. In the United States during the 1960s, the various state governments had vastly different rules about advertising by optometrists. Some states allowed advertising for eyeglasses and eye examinations. Many states, however, prohibited it. For example, the Florida law read as follows:

It is unlawful for any person, firm, or corporation to . . . advertise either directly or indirectly by any means whatsoever any definite or indefinite price or credit terms on prescriptive or corrective lens, frames, complete prescriptive or corrective glasses, or any optometric service. . . . This section is passed in the interest of public health, safety, and welfare, and its provisions shall be liberally construed to carry out its objects and purposes.

Professional optometrists enthusiastically endorsed these restrictions on advertising.

Benham used the differences in state law as a natural experiment to test the two views of advertising. The results were striking. In those states that prohibited advertising, the average price paid for a pair of eyeglasses was \$33. (This number is not as low as it seems, for this price is from 1963, when all prices were much lower than they are today. To convert 1963 prices into today’s dollars, you can multiply them by 5.) In those states that did not restrict

advertising, the average price was \$26. Thus, advertising reduced average prices by more than 20 percent. In the market for eyeglasses, and probably in many other markets as well, advertising fosters competition and leads to lower prices for consumers.

ADVERTISING AS A SIGNAL OF QUALITY

Many types of advertising contain little apparent information about the product being advertised. Consider a firm introducing a new breakfast cereal. A typical advertisement might have some highly paid actor eating the cereal and exclaiming how wonderful it tastes. How much information does the advertisement really provide?

The answer is: more than you might think. Defenders of advertising argue that even advertising that appears to contain little hard information may in fact tell consumers something about product quality. The willingness of the firm to spend a large amount of money on advertising can itself be a *signal* to consumers about the quality of the product being offered.

Consider the problem facing two firms—Post and Kellogg. Each company has just come up with a recipe for a new cereal, which it would sell for \$3 a box. To keep things simple, let's assume that the marginal cost of making cereal is zero, so the \$3 is all profit. Each company knows that if it spends \$10 million on advertising, it will get 1 million consumers to try its new cereal. And each company knows that if consumers like the cereal, they will buy it not once but many times.

First consider Post's decision. Based on market research, Post knows that its cereal is only mediocre. Although advertising would sell one box to each of 1 million consumers, the consumers would quickly learn that the cereal is not very good and stop buying it. Post decides it is not worth paying \$10 million in advertising to get only \$3 million in sales. So it does not bother to advertise. It sends its cooks back to the drawing board to find another recipe.

Kellogg, on the other hand, knows that its cereal is great. Each person who tries it will buy a box a month for the next year. Thus, the \$10 million in advertising will bring in \$36 million in sales. Advertising is profitable here because Kellogg has a good product that consumers will buy repeatedly. Thus, Kellogg chooses to advertise.

Now that we have considered the behavior of the two firms, let's consider the behavior of consumers. We began by asserting that consumers are inclined to try a new cereal that they see advertised. But is this behavior rational? Should a consumer try a new cereal just because the seller has chosen to advertise it?

In fact, it may be completely rational for consumers to try new products that they see advertised. In our story, consumers decide to try Kellogg's new cereal because Kellogg advertises. Kellogg chooses to advertise because it knows that its cereal is quite good, while Post chooses not to advertise because it knows that its cereal is only mediocre. By its willingness to spend money on advertising, Kellogg signals to consumers the quality of its cereal. Each consumer thinks, quite sensibly, "Boy, if the Kellogg Company is willing to spend so much money advertising this new cereal, it must be really good."

What is most surprising about this theory of advertising is that the content of the advertisement is irrelevant. Kellogg signals the quality of its product by its willingness to spend money on advertising. What the advertisements say is not as

important as the fact that consumers know ads are expensive. By contrast, cheap advertising cannot be effective at signaling quality to consumers. In our example, if an advertising campaign cost less than \$3 million, both Post and Kellogg would use it to market their new cereals. Because both good and mediocre cereals would be advertised, consumers could not infer the quality of a new cereal from the fact that it is advertised. Over time, consumers would learn to ignore such cheap advertising.

This theory can explain why firms pay famous actors large amounts of money to make advertisements that, on the surface, appear to convey no information at all. The information is not in the advertisement's content, but simply in its existence and expense.

BRAND NAMES

Advertising is closely related to the existence of brand names. In many markets, there are two types of firms. Some firms sell products with widely recognized brand names, while other firms sell generic substitutes. For example, in a typical drugstore, you can find Bayer aspirin on the shelf next to a generic aspirin. In a typical grocery store, you can find Pepsi next to less familiar colas. Most often, the firm with the brand name spends more on advertising and charges a higher price for its product.

Just as there is disagreement about the economics of advertising, there is disagreement about the economics of brand names. Let's consider both sides of the debate.

Critics of brand names argue that brand names cause consumers to perceive differences that do not really exist. In many cases, the generic good is almost indistinguishable from the brand-name good. Consumers' willingness to pay more for the brand-name good, these critics assert, is a form of irrationality fostered by advertising. Economist Edward Chamberlin, one of the early developers of the theory of monopolistic competition, concluded from this argument that brand names were bad for the economy. He proposed that the government discourage



their use by refusing to enforce the exclusive trademarks that companies use to identify their products.

More recently, economists have defended brand names as a useful way for consumers to ensure that the goods they buy are of high quality. There are two related arguments. First, brand names provide consumers *information* about quality when quality cannot be easily judged in advance of purchase. Second, brand names give firms an *incentive* to maintain high quality, because firms have a financial stake in maintaining the reputation of their brand names.

To see how these arguments work in practice, consider a famous brand name: McDonald's hamburgers. Imagine that you are driving through an unfamiliar town and want to stop for lunch. You see a McDonald's and a local restaurant next to it. Which do you choose? The local restaurant may in fact offer better food at lower prices, but you have no way of knowing that. By contrast, McDonald's offers a consistent product across many cities. Its brand name is useful to you as a way of judging the quality of what you are about to buy.

The McDonald's brand name also ensures that the company has an incentive to maintain quality. For example, if some customers were to become ill from bad food sold at a McDonald's, the news would be disastrous for the company. McDonald's would lose much of the valuable reputation that it has built up with years of expensive advertising. As a result, it would lose sales and profit not just in the outlet that sold the bad food but in its many outlets throughout the country. By contrast, if some customers were to become ill from bad food at a local restaurant, that restaurant might have to close down, but the lost profits would be much smaller. Hence, McDonald's has a greater incentive to ensure that its food is safe.

The debate over brand names thus centers on the question of whether consumers are rational in preferring brand names over generic substitutes. Critics of brand names argue that brand names are the result of an irrational consumer response to advertising. Defenders of brand names argue that consumers have good reason to pay more for brand-name products because they can be more confident in the quality of these products.

CASE STUDY BRAND NAMES UNDER COMMUNISM

Defenders of brand names get some support for their view from experiences in the former Soviet Union. When the Soviet Union adhered to the principles of communism, central planners in the government replaced the invisible hand of the marketplace. Yet, just like consumers living in an economy with free markets, Soviet central planners learned that brand names were useful in helping to ensure product quality.

In an article published in the *Journal of Political Economy* in 1960, Marshall Goldman, an expert on the Soviet economy, described the Soviet experience:

In the Soviet Union, production goals have been set almost solely in quantitative or value terms, with the result that, in order to meet the plan, quality is often sacrificed. . . . Among the methods adopted by the Soviets to deal with this problem, one is of particular interest to us—intentional product differentiation. . . . In order to distinguish one firm from similar firms in the same industry or ministry, each firm has its own name. Whenever it is

IN THE NEWS*TV Networks as
Brand Names*

BRAND NAMES CONVEY INFORMATION TO consumers about the goods that firms are offering. Establishing a brand name—and ensuring that it conveys the right information—is an important strategy for many businesses, including TV networks.

**A TV Season When
Image Is Everything**

BY STUART ELLIOTT

A marketing blitz to promote fall television programming, estimated at a record \$400 million to \$500 million, has been inundating America with a barrage of branding.

Branding is a shorthand term along Madison Avenue for attempts to create or burnish an identity or image, just as Coca-Cola seeks to distinguish itself from Pepsi-Cola. For the 1996–97 prime-time broadcast television season,

which officially began this week, viewers have been swamped by the torrent of teasing practically since the 1995–96 season ended in May.

At the center of those efforts is the most ambitious push ever by the broadcast networks to brand themselves and many of the blocks of programming they offer—a marked departure from the past, when they would promote only specific shows.

“The perception was that people watched shows, not networks,” said Bob Bibb, who with Lewis Goldstein jointly heads marketing for WB, a fledgling network owned by Time Warner, Inc., and based in Burbank, California.

“But that was when there were only three networks, three choices,” Mr. Bibb added, “and it was easy to find the shows you liked.”

WB has been presenting a sassy singing cartoon character named Michigan J. Frog as its “spokesphibian,” personifying the entire lineup of the “Dubba-dubba-WB”—as he insists upon calling the network.

“It’s not a frog, it’s an attitude,” Mr. Bibb said, “a consistency from show to show.”

In television, an intrinsic part of branding is selecting shows that seem related and might appeal to a certain audience segment. It means “developing



AN ATTITUDE, NOT JUST A FROG

an overall packaging of the network to build a relationship with viewers, so they will come to expect certain things from us,” said Alan Cohen, executive vice president for the ABC-TV unit of the Walt Disney Company in New York.

That, he said, means defining the network so that “when you’re watching ABC, you’ll *know* you’re watching ABC”—and to accomplish it in a way that appeals to the primary ABC audience of youngish urbanites and families with children.

SOURCE: *The New York Times*, September 20, 1996, p. D1.

physically possible, it is obligatory that the firm identify itself on the good or packaging with a “production mark.”

Goldman quotes the analysis of a Soviet marketing expert:

This [trademark] makes it easy to establish the actual producer of the product in case it is necessary to call him to account for the poor quality of his goods. For this reason, it is one of the most effective weapons in the battle for the quality of products. . . . The trademark makes it possible for the consumer to select the good which he likes. . . . This forces other firms to undertake measures to improve the quality of their own product in harmony with the demands of the consumer.

Goldman notes that “these arguments are clear enough and sound as if they might have been written by a bourgeois apologist.”

QUICK QUIZ: How might advertising make markets less competitive? How might it make markets more competitive? ♦ Give the arguments for and against brand names.

CONCLUSION

Monopolistic competition is true to its name: It is a hybrid of monopoly and competition. Like a monopoly, each monopolistic competitor faces a downward-sloping demand curve and, as a result, charges a price above marginal cost. As in a competitive market, however, there are many firms, and entry and exit drive the profit of each monopolistic competitor toward zero. Because monopolistically competitive firms produce differentiated products, each firm advertises in order to attract customers to its own brand. To some extent, advertising manipulates consumers' tastes, promotes irrational brand loyalty, and impedes competition. To a larger extent, advertising provides information, establishes brand names of reliable quality, and fosters competition.

The theory of monopolistic competition seems to describe many markets in the economy. It is somewhat disappointing, therefore, that the theory does not yield simple and compelling advice for public policy. From the standpoint of the economic theorist, the allocation of resources in monopolistically competitive markets is not perfect. Yet, from the standpoint of a practical policymaker, there may be little that can be done to improve it.

Summary

- ♦ A monopolistically competitive market is characterized by three attributes: many firms, differentiated products, and free entry.
- ♦ The equilibrium in a monopolistically competitive market differs from that in a perfectly competitive market in two related ways. First, each firm has excess capacity. That is, it operates on the downward-sloping portion of the average-total-cost curve. Second, each firm charges a price above marginal cost.
- ♦ Monopolistic competition does not have all the desirable properties of perfect competition. There is the standard deadweight loss of monopoly caused by the markup of price over marginal cost. In addition, the number of firms (and thus the variety of products) can be too large or too small. In practice, the ability of policymakers to correct these inefficiencies is limited.
- ♦ The product differentiation inherent in monopolistic competition leads to the use of advertising and brand names. Critics of advertising and brand names argue that firms use them to take advantage of consumer irrationality and to reduce competition. Defenders of advertising and brand names argue that firms use them to inform consumers and to compete more vigorously on price and product quality.

Key Concepts

monopolistic competition, p. 378

Questions for Review

- Describe the three attributes of monopolistic competition. How is monopolistic competition like monopoly? How is it like perfect competition?
- Draw a diagram depicting a firm in a monopolistically competitive market that is making profits. Now show what happens to this firm as new firms enter the industry.
- Draw a diagram of the long-run equilibrium in a monopolistically competitive market. How is price related to average total cost? How is price related to marginal cost?
- Does a monopolistic competitor produce too much or too little output compared to the most efficient level? What practical considerations make it difficult for policymakers to solve this problem?
- How might advertising reduce economic well-being? How might advertising increase economic well-being?
- How might advertising with no apparent informational content in fact convey information to consumers?
- Explain two benefits that might arise from the existence of brand names.

Problems and Applications

- Classify the following markets as perfectly competitive, monopolistic, or monopolistically competitive, and explain your answers.
 - wooden #2 pencils
 - bottled water
 - copper
 - local telephone service
 - peanut butter
 - lipstick
- What feature of the product being sold distinguishes a monopolistically competitive firm from a monopolistic firm?
- The chapter states that monopolistically competitive firms could increase the quantity they produce and lower the average total cost of production. Why don't they do so?
- Sparkle is one firm of many in the market for toothpaste, which is in long-run equilibrium.
 - Draw a diagram showing Sparkle's demand curve, marginal-revenue curve, average-total-cost curve, and marginal-cost curve. Label Sparkle's profit-maximizing output and price.
 - What is Sparkle's profit? Explain.
 - On your diagram, show the consumer surplus derived from the purchase of Sparkle toothpaste. Also show the deadweight loss relative to the efficient level of output.
 - If the government forced Sparkle to produce the efficient level of output, what would happen to the firm? What would happen to Sparkle's customers?
- Do monopolistically competitive markets typically have the optimal number of products? Explain.
- Complete the table below by filling in YES, NO, or MAYBE for each type of market structure.

Do Firms:	PERFECT COMPETITION	MONOPOLISTIC COMPETITION	MONOPOLY
Make differentiated products?	_____	_____	_____
Have excess capacity?	_____	_____	_____
Advertise?	_____	_____	_____
Pick Q so that $MR = MC$?	_____	_____	_____
Pick Q so that $P = MC$?	_____	_____	_____
Earn economic profits in long-run equilibrium?	_____	_____	_____
Face a downward-sloping demand curve?	_____	_____	_____
Have MR less than price?	_____	_____	_____
Face the entry of other firms?	_____	_____	_____
Exit in the long run if profits are less than zero?	_____	_____	_____

7. The chapter says that monopolistically competitive firms may send Christmas cards to their customers. What do they accomplish by this? Explain in words and with a diagram.
8. If you were thinking of entering the ice-cream business, would you try to make ice cream that is just like one of the existing brands? Explain your decision using the ideas of this chapter.
9. Describe three commercials that you have seen on TV. In what ways, if any, were each of these commercials socially useful? In what ways were they socially wasteful? Did the commercials affect the likelihood of your buying the product, and why?
10. For each of the following pairs of firms, explain which firm would be more likely to engage in advertising.
 - a. a family-owned farm or a family-owned restaurant
 - b. a manufacturer of forklifts or a manufacturer of cars
 - c. a company that invented a very reliable watch or a company that invented a less reliable watch that costs the same amount to make
11. Twenty years ago the market for chicken was perfectly competitive. Then Frank Perdue began marketing chicken under his name.
 - a. How do you suppose Perdue created a brand name for chicken? What did he gain from doing so?
 - b. What did society gain from having brand-name chicken? What did society lose?
12. The makers of Tylenol pain reliever do a lot of advertising and have very loyal customers. In contrast, the makers of generic acetaminophen do no advertising, and their customers shop only for the lowest price. Assume that the marginal costs of Tylenol and generic acetaminophen are the same and constant.
 - a. Draw a diagram showing Tylenol's demand, marginal-revenue, and marginal-cost curves. Label Tylenol's price and markup over marginal cost.
 - b. Repeat part (a) for a producer of generic acetaminophen. How do the diagrams differ? Which company has the bigger markup? Explain.
 - c. Which company has the bigger incentive for careful quality control? Why?

18



THE MARKETS FOR THE FACTORS OF PRODUCTION

When you finish school, your income will be determined largely by what kind of job you take. If you become a computer programmer, you will earn more than if you become a gas station attendant. This fact is not surprising, but it is not obvious why it is true. No law requires that computer programmers be paid more than gas station attendants. No ethical principle says that programmers are more deserving. What then determines which job will pay you the higher wage?

Your income, of course, is a small piece of a larger economic picture. In 1999 the total income of all U.S. residents was about \$8 trillion. People earned this income in various ways. Workers earned about three-fourths of it in the form of wages and fringe benefits. The rest went to landowners and to the owners of *capital*—the economy's stock of equipment and structures—in the form of rent, profit, and interest. What determines how much goes to workers? To landowners? To the owners of capital? Why do some workers earn higher wages than others, some

IN THIS CHAPTER YOU WILL . . .

Analyze the labor demand of competitive, profit-maximizing firms

Consider the household decisions that lie behind labor supply

Learn why equilibrium wages equal the value of the marginal product of labor

Consider how the other factors of production—land and capital—are compensated

Examine how a change in the supply of one factor alters the earnings of all the factors

landowners higher rental income than others, and some capital owners greater profit than others? Why, in particular, do computer programmers earn more than gas station attendants?

The answers to these questions, like most in economics, hinge on supply and demand. The supply and demand for labor, land, and capital determine the prices paid to workers, landowners, and capital owners. To understand why some people have higher incomes than others, therefore, we need to look more deeply at the markets for the services they provide. That is our job in this and the next two chapters.

factors of production

the inputs used to produce goods and services

This chapter provides the basic theory for the analysis of factor markets. As you may recall from Chapter 2, the **factors of production** are the inputs used to produce goods and services. Labor, land, and capital are the three most important factors of production. When a computer firm produces a new software program, it uses programmers' time (labor), the physical space on which its offices sit (land), and an office building and computer equipment (capital). Similarly, when a gas station sells gas, it uses attendants' time (labor), the physical space (land), and the gas tanks and pumps (capital).

Although in many ways factor markets resemble the goods markets we have analyzed in previous chapters, they are different in one important way: The demand for a factor of production is a *derived demand*. That is, a firm's demand for a factor of production is derived from its decision to supply a good in another market. The demand for computer programmers is inextricably tied to the supply of computer software, and the demand for gas station attendants is inextricably tied to the supply of gasoline.

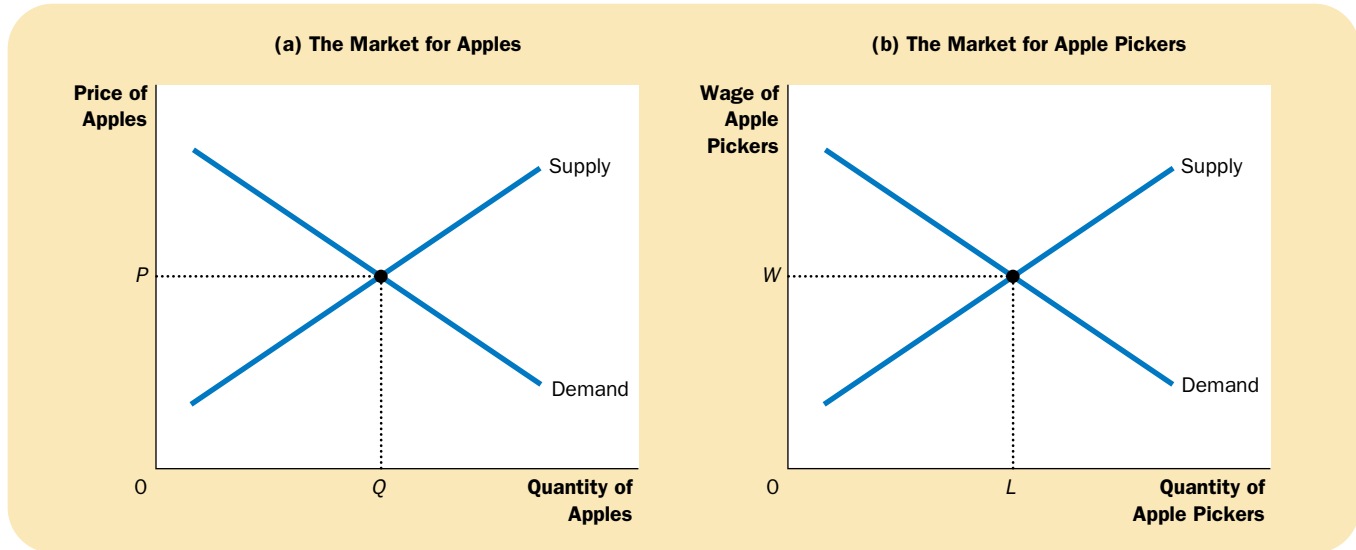
In this chapter we analyze factor demand by considering how a competitive, profit-maximizing firm decides how much of any factor to buy. We begin our analysis by examining the demand for labor. Labor is the most important factor of production, for workers receive most of the total income earned in the U.S. economy. Later in the chapter, we see that the lessons we learn about the labor market apply directly to the markets for the other factors of production.

The basic theory of factor markets developed in this chapter takes a large step toward explaining how the income of the U.S. economy is distributed among workers, landowners, and owners of capital. Chapter 19 will build on this analysis to examine in more detail why some workers earn more than others. Chapter 20 will examine how much inequality results from this process and then consider what role the government should and does play in altering the distribution of income.

THE DEMAND FOR LABOR

Labor markets, like other markets in the economy, are governed by the forces of supply and demand. This is illustrated in Figure 18-1. In panel (a) the supply and demand for apples determine the price of apples. In panel (b) the supply and demand for apple pickers determine the price, or wage, of apple pickers.

As we have already noted, labor markets are different from most other markets because labor demand is a derived demand. Most labor services, rather than



THE VERSATILITY OF SUPPLY AND DEMAND. The basic tools of supply and demand apply to goods and to labor services. Panel (a) shows how the supply and demand for apples determine the price of apples. Panel (b) shows how the supply and demand for apple pickers determine the wage of apple pickers.

Figure 18-1

being final goods ready to be enjoyed by consumers, are inputs into the production of other goods. To understand labor demand, we need to focus on the firms that hire the labor and use it to produce goods for sale. By examining the link between the production of goods and the demand for labor, we gain insight into the determination of equilibrium wages.

THE COMPETITIVE PROFIT-MAXIMIZING FIRM

Let's look at how a typical firm, such as an apple producer, decides the quantity of labor to demand. The firm owns an apple orchard and each week must decide how many apple pickers to hire to harvest its crop. After the firm makes its hiring decision, the workers pick as many apples as they can. The firm then sells the apples, pays the workers, and keeps what is left as profit.

We make two assumptions about our firm. First, we assume that our firm is *competitive* both in the market for apples (where the firm is a seller) and in the market for apple pickers (where the firm is a buyer). Recall from Chapter 14 that a competitive firm is a price taker. Because there are many other firms selling apples and hiring apple pickers, a single firm has little influence over the price it gets for apples or the wage it pays apple pickers. The firm takes the price and the wage as given by market conditions. It only has to decide how many workers to hire and how many apples to sell.

Second, we assume that the firm is *profit-maximizing*. Thus, the firm does not directly care about the number of workers it has or the number of apples it produces. It cares only about profit, which equals the total revenue from the sale of

LABOR	OUTPUT	MARGINAL PRODUCT OF LABOR	VALUE OF THE MARGINAL PRODUCT OF LABOR	WAGE	MARGINAL PROFIT
L (NUMBER OF WORKERS)	Q (BUSHELS PER WEEK)	$MPL = \Delta Q / \Delta L$ (BUSHELS PER WEEK)	$VMPL = P \times MPL$	W	$\Delta \text{PROFIT} =$ $VMPL - W$
0	0	100	\$1,000	\$500	\$500
1	100	80	800	500	300
2	180	60	600	500	100
3	240	40	400	500	-100
4	280	20	200	500	-300
5	300				

Table 18-1

HOW THE COMPETITIVE FIRM DECIDES HOW MUCH LABOR TO HIRE

apples minus the total cost of producing them. The firm's supply of apples and its demand for workers are derived from its primary goal of maximizing profit.

THE PRODUCTION FUNCTION AND THE MARGINAL PRODUCT OF LABOR

To make its hiring decision, the firm must consider how the size of its workforce affects the amount of output produced. In other words, it must consider how the number of apple pickers affects the quantity of apples it can harvest and sell. Table 18-1 gives a numerical example. In the first column is the number of workers. In the second column is the quantity of apples the workers harvest each week.

These two columns of numbers describe the firm's ability to produce. As we noted in Chapter 13, economists use the term **production function** to describe the relationship between the quantity of the inputs used in production and the quantity of output from production. Here the "input" is the apple pickers and the "output" is the apples. The other inputs—the trees themselves, the land, the firm's trucks and tractors, and so on—are held fixed for now. This firm's production function shows that if the firm hires 1 worker, that worker will pick 100 bushels of apples per week. If the firm hires 2 workers, the two workers together will pick 180 bushels per week, and so on.

Figure 18-2 graphs the data on labor and output presented in Table 18-1. The number of workers is on the horizontal axis, and the amount of output is on the vertical axis. This figure illustrates the production function.

One of the *Ten Principles of Economics* introduced in Chapter 1 is that rational people think at the margin. This idea is the key to understanding how firms decide what quantity of labor to hire. To take a step toward this decision, the third column in Table 18-1 gives the **marginal product of labor**, the increase in the amount of output from an additional unit of labor. When the firm increases the number of workers from 1 to 2, for example, the amount of apples produced rises from 100 to 180 bushels. Therefore, the marginal product of the second worker is 80 bushels.

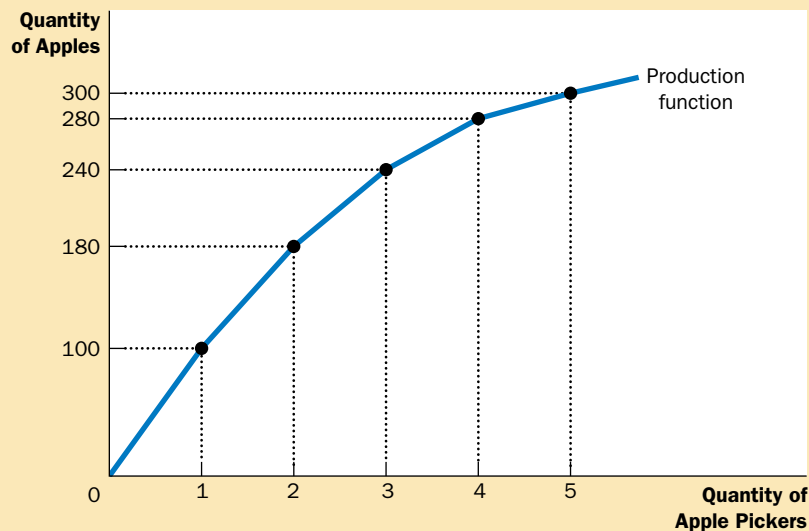
production function

the relationship between the quantity of inputs used to make a good and the quantity of output of that good



marginal product of labor

the increase in the amount of output from an additional unit of labor

**Figure 18-2**

THE PRODUCTION FUNCTION. The production function is the relationship between the inputs into production (apple pickers) and the output from production (apples). As the quantity of the input increases, the production function gets flatter, reflecting the property of diminishing marginal product.

Notice that as the number of workers increases, the marginal product of labor declines. As you may recall from Chapter 13, this property is called **diminishing marginal product**. At first, when only a few workers are hired, they pick apples from the best trees in the orchard. As the number of workers increases, additional workers have to pick from the trees with fewer apples. Hence, as more and more workers are hired, each additional worker contributes less to the production of apples. For this reason, the production function in Figure 18-2 becomes flatter as the number of workers rises.

diminishing marginal product

the property whereby the marginal product of an input declines as the quantity of the input increases

THE VALUE OF THE MARGINAL PRODUCT AND THE DEMAND FOR LABOR

Our profit-maximizing firm is concerned more with money than with apples. As a result, when deciding how many workers to hire, the firm considers how much profit each worker would bring in. Because profit is total revenue minus total cost, the profit from an additional worker is the worker's contribution to revenue minus the worker's wage.

To find the worker's contribution to revenue, we must convert the marginal product of labor (which is measured in bushels of apples) into the *value* of the marginal product (which is measured in dollars). We do this using the price of apples. To continue our example, if a bushel of apples sells for \$10 and if an additional worker produces 80 bushels of apples, then the worker produces \$800 of revenue.

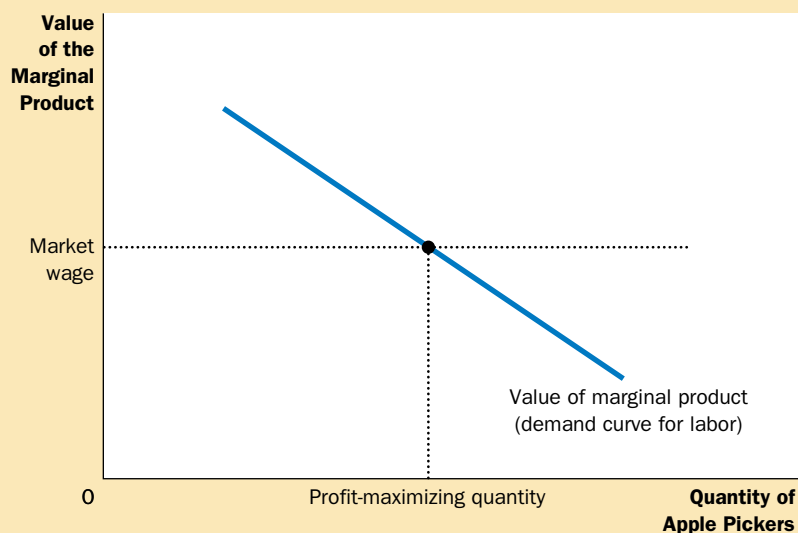
The **value of the marginal product** of any input is the marginal product of that input multiplied by the market price of the output. The fourth column in Table 18-1 shows the value of the marginal product of labor in our example, assuming the price of apples is \$10 per bushel. Because the market price is constant for a competitive firm, the value of the marginal product (like the marginal product itself) diminishes as the number of workers rises.

value of the marginal product

the marginal product of an input times the price of the output

Figure 18-3

THE VALUE OF THE MARGINAL PRODUCT OF LABOR. This figure shows how the value of the marginal product (the marginal product times the price of the output) depends on the number of workers. The curve slopes downward because of diminishing marginal product. For a competitive, profit-maximizing firm, this value-of-marginal-product curve is also the firm's labor demand curve.



Now consider how many workers the firm will hire. Suppose that the market wage for apple pickers is \$500 per week. In this case, the first worker that the firm hires is profitable: The first worker yields \$1,000 in revenue, or \$500 in profit. Similarly, the second worker yields \$800 in additional revenue, or \$300 in profit. The third worker produces \$600 in additional revenue, or \$100 in profit. After the third worker, however, hiring workers is unprofitable. The fourth worker would yield only \$400 of additional revenue. Because the worker's wage is \$500, hiring the fourth worker would mean a \$100 reduction in profit. Thus, the firm hires only three workers.

It is instructive to consider the firm's decision graphically. Figure 18-3 graphs the value of the marginal product. This curve slopes downward because the marginal product of labor diminishes as the number of workers rises. The figure also includes a horizontal line at the market wage. To maximize profit, the firm hires workers up to the point where these two curves cross. Below this level of employment, the value of the marginal product exceeds the wage, so hiring another worker would increase profit. Above this level of employment, the value of the marginal product is less than the wage, so the marginal worker is unprofitable. Thus, *a competitive, profit-maximizing firm hires workers up to the point where the value of the marginal product of labor equals the wage.*

Having explained the profit-maximizing hiring strategy for a competitive firm, we can now offer a theory of labor demand. Recall that a firm's labor demand curve tells us the quantity of labor that a firm demands at any given wage. We have just seen in Figure 18-3 that the firm makes that decision by choosing the quantity of labor at which the value of the marginal product equals the wage. As a result, *the value-of-marginal-product curve is the labor demand curve for a competitive, profit-maximizing firm.*

FYI*Input Demand and Output Supply: Two Sides of the Same Coin*

function links the quantity of inputs to the quantity of output, you should not be surprised to learn that the firm's decision about input demand is closely linked to its decision about output supply. In fact, these two decisions are two sides of the same coin.

To see this relationship more fully, let's consider how the marginal product of labor (MPL) and marginal cost (MC) are related. Suppose an additional worker costs \$500 and has a marginal product of 50 bushels of apples. In this case, producing 50 more bushels costs \$500; the marginal cost of a bushel is \$500/50, or \$10. More generally, if W is the wage, and an extra unit of labor produces MPL units of output, then the marginal cost of a unit of output is $MC = W/MPL$.

This analysis shows that diminishing marginal product is closely related to increasing marginal cost. When our apple orchard grows crowded with workers, each additional

worker adds less to the production of apples (MPL falls). Similarly, when the apple firm is producing a large quantity of apples, the orchard is already crowded with workers, so it is more costly to produce an additional bushel of apples (MC rises).

Now consider our criterion for profit maximization. We determined earlier that a profit-maximizing firm chooses the quantity of labor so that the value of the marginal product ($P \times MPL$) equals the wage (W). We can write this mathematically as

$$P \times MPL = W.$$

If we divide both sides of this equation by MPL , we obtain

$$P = W/MPL.$$

We just noted that W/MPL equals marginal cost MC . Therefore, we can substitute to obtain

$$P = MC.$$

This equation states that the price of the firm's output is equal to the marginal cost of producing a unit of output. *Thus, when a competitive firm hires labor up to the point at which the value of the marginal product equals the wage, it also produces up to the point at which the price equals marginal cost.* Our analysis of labor demand in this chapter is just another way of looking at the production decision we first saw in Chapter 14.

WHAT CAUSES THE LABOR DEMAND CURVE TO SHIFT?

We now understand the labor demand curve: It is nothing more than a reflection of the value of marginal product of labor. With this insight in mind, let's consider a few of the things that might cause the labor demand curve to shift.

The Output Price The value of the marginal product is marginal product times the price of the firm's output. Thus, when the output price changes, the value of the marginal product changes, and the labor demand curve shifts. An increase in the price of apples, for instance, raises the value of the marginal product of each worker that picks apples and, therefore, increases labor demand from the firms that supply apples. Conversely, a decrease in the price of apples reduces the value of the marginal product and decreases labor demand.

Technological Change Between 1968 and 1998, the amount of output a typical U.S. worker produced in an hour rose by 57 percent. Why? The most

important reason is technological progress: Scientists and engineers are constantly figuring out new and better ways of doing things. This has profound implications for the labor market. Technological advance raises the marginal product of labor, which in turn increases the demand for labor. Such technological advance explains persistently rising employment in face of rising wages: Even though wages (adjusted for inflation) increased by 62 percent over these three decades, firms nonetheless increased by 72 percent the number of workers they employed.

The Supply of Other Factors The quantity available of one factor of production can affect the marginal product of other factors. A fall in the supply of ladders, for instance, will reduce the marginal product of apple pickers and thus the demand for apple pickers. We consider this linkage among the factors of production more fully later in the chapter.

QUICK QUIZ: Define *marginal product of labor* and *value of the marginal product of labor*. ♦ Describe how a competitive, profit-maximizing firm decides how many workers to hire.

THE SUPPLY OF LABOR

Having analyzed labor demand in detail, let's turn to the other side of the market and consider labor supply. A formal model of labor supply is included in Chapter 21, where we develop the theory of household decisionmaking. Here we discuss briefly and informally the decisions that lie behind the labor supply curve.

THE TRADEOFF BETWEEN WORK AND LEISURE



One of the *Ten Principles of Economics* in Chapter 1 is that people face tradeoffs. Probably no tradeoff is more obvious or more important in a person's life than the tradeoff between work and leisure. The more hours you spend working, the fewer hours you have to watch TV, have dinner with friends, or pursue your favorite hobby. The tradeoff between labor and leisure lies behind the labor supply curve.



Another one of the *Ten Principles of Economics* is that the cost of something is what you give up to get it. What do you give up to get an hour of leisure? You give up an hour of work, which in turn means an hour of wages. Thus, if your wage is \$15 per hour, the opportunity cost of an hour of leisure is \$15. And when you get a raise to \$20 per hour, the opportunity cost of enjoying leisure goes up.

The labor supply curve reflects how workers' decisions about the labor-leisure tradeoff respond to a change in that opportunity cost. An upward-sloping labor supply curve means that an increase in the wage induces workers to increase the quantity of labor they supply. Because time is limited, more hours of work means that workers are enjoying less leisure. That is, workers respond to the increase in the opportunity cost of leisure by taking less of it.

It is worth noting that the labor supply curve need not be upward sloping. Imagine you got that raise from \$15 to \$20 per hour. The opportunity cost of

leisure is now greater, but you are also richer than you were before. You might decide that with your extra wealth you can now afford to enjoy more leisure; in this case, your labor supply curve would slope backwards. In Chapter 21, we discuss this possibility in terms of conflicting effects on your labor-supply decision (called income and substitution effects). For now, we ignore the possibility of backward-sloping labor supply and assume that the labor supply curve is upward sloping.

WHAT CAUSES THE LABOR SUPPLY CURVE TO SHIFT?

The labor supply curve shifts whenever people change the amount they want to work at a given wage. Let's now consider some of the events that might cause such a shift.

Changes in Tastes In 1950, 34 percent of women were employed at paid jobs or looking for work. In 1998, the number had risen to 60 percent. There are, of course, many explanations for this development, but one of them is changing tastes, or attitudes toward work. A generation or two ago, it was the norm for women to stay at home while raising children. Today, family sizes are smaller, and more mothers choose to work. The result is an increase in the supply of labor.

Changes in Alternative Opportunities The supply of labor in any one labor market depends on the opportunities available in other labor markets. If the wage earned by pear pickers suddenly rises, some apple pickers may choose to switch occupations. The supply of labor in the market for apple pickers falls.

Immigration Movements of workers from region to region, or country to country, is an obvious and often important source of shifts in labor supply. When immigrants come to the United States, for instance, the supply of labor in the United States increases and the supply of labor in the immigrants' home countries contracts. In fact, much of the policy debate about immigration centers on its effect on labor supply and, thereby, equilibrium in the labor market.

QUICK QUIZ: Who has a greater opportunity cost of enjoying leisure—a janitor or a brain surgeon? Explain. Can this help explain why doctors work such long hours?

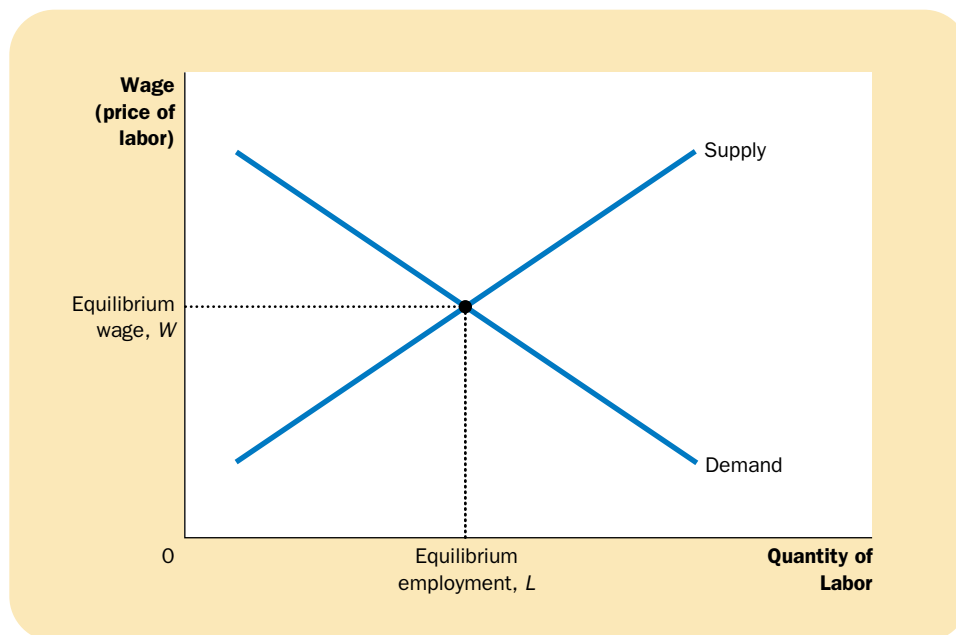
EQUILIBRIUM IN THE LABOR MARKET

So far we have established two facts about how wages are determined in competitive labor markets:

- ◆ The wage adjusts to balance the supply and demand for labor.
- ◆ The wage equals the value of the marginal product of labor.

Figure 18-4

EQUILIBRIUM IN A LABOR MARKET. Like all prices, the price of labor (the wage) depends on supply and demand. Because the demand curve reflects the value of the marginal product of labor, in equilibrium workers receive the value of their marginal contribution to the production of goods and services.



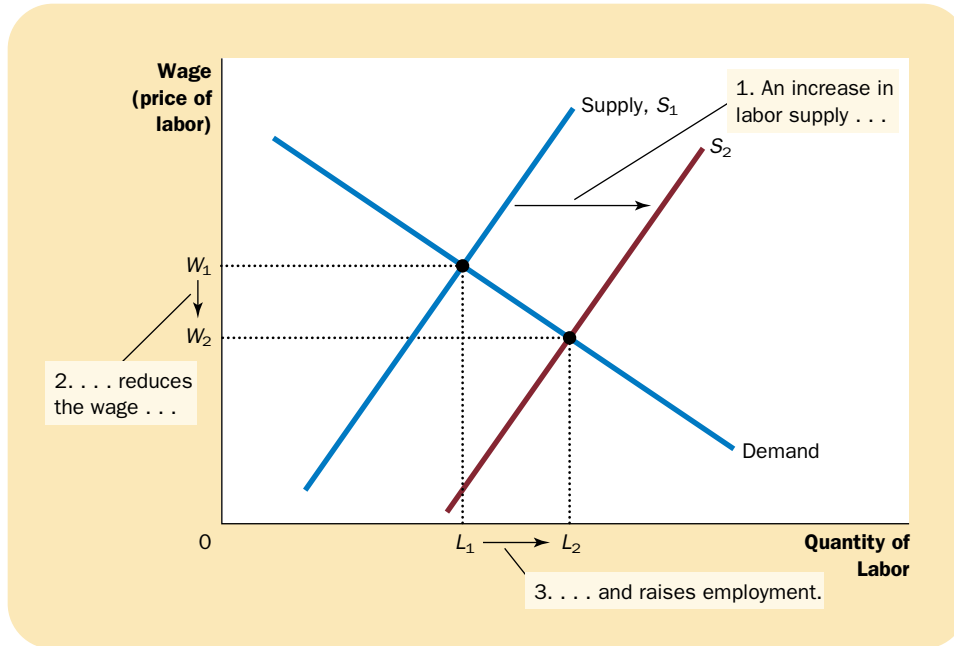
At first, it might seem surprising that the wage can do both these things at once. In fact, there is no real puzzle here, but understanding why there is no puzzle is an important step to understanding wage determination.

Figure 18-4 shows the labor market in equilibrium. The wage and the quantity of labor have adjusted to balance supply and demand. When the market is in this equilibrium, each firm has bought as much labor as it finds profitable at the equilibrium wage. That is, each firm has followed the rule for profit maximization: It has hired workers until the value of the marginal product equals the wage. Hence, the wage must equal the value of marginal product of labor once it has brought supply and demand into equilibrium.

This brings us to an important lesson: *Any event that changes the supply or demand for labor must change the equilibrium wage and the value of the marginal product by the same amount, because these must always be equal.* To see how this works, let's consider some events that shift these curves.

SHIFTS IN LABOR SUPPLY

Suppose that immigration increases the number of workers willing to pick apples. As Figure 18-5 shows, the supply of labor shifts to the right from S_1 to S_2 . At the initial wage W_1 , the quantity of labor supplied now exceeds the quantity demanded. This surplus of labor puts downward pressure on the wage of apple pickers, and the fall in the wage from W_1 to W_2 in turn makes it profitable for firms to hire more workers. As the number of workers employed in each apple orchard rises, the marginal product of a worker falls, and so does the value of the marginal product. In the new equilibrium, both the wage and the value of the marginal product of labor are lower than they were before the influx of new workers.

**Figure 18-5****A SHIFT IN LABOR SUPPLY.**

When labor supply increases from S_1 to S_2 , perhaps because of an immigration of new workers, the equilibrium wage falls from W_1 to W_2 . At this lower wage, firms hire more labor, so employment rises from L_1 to L_2 . The change in the wage reflects a change in the value of the marginal product of labor: With more workers, the added output from an extra worker is smaller.

An episode from Israel illustrates how a shift in labor supply can alter the equilibrium in a labor market. During most of the 1980s, many thousands of Palestinians regularly commuted from their homes in the Israeli-occupied West Bank and Gaza Strip to jobs in Israel, primarily in the construction and agriculture industries. In 1988, however, political unrest in these occupied areas induced the Israeli government to take steps that, as a by-product, reduced this supply of workers. Curfews were imposed, work permits were checked more thoroughly, and a ban on overnight stays of Palestinians in Israel was enforced more rigorously. The economic impact of these steps was exactly as theory predicts: The number of Palestinians with jobs in Israel fell by half, while those who continued to work in Israel enjoyed wage increases of about 50 percent. With a reduced number of Palestinian workers in Israel, the value of the marginal product of the remaining workers was much higher.

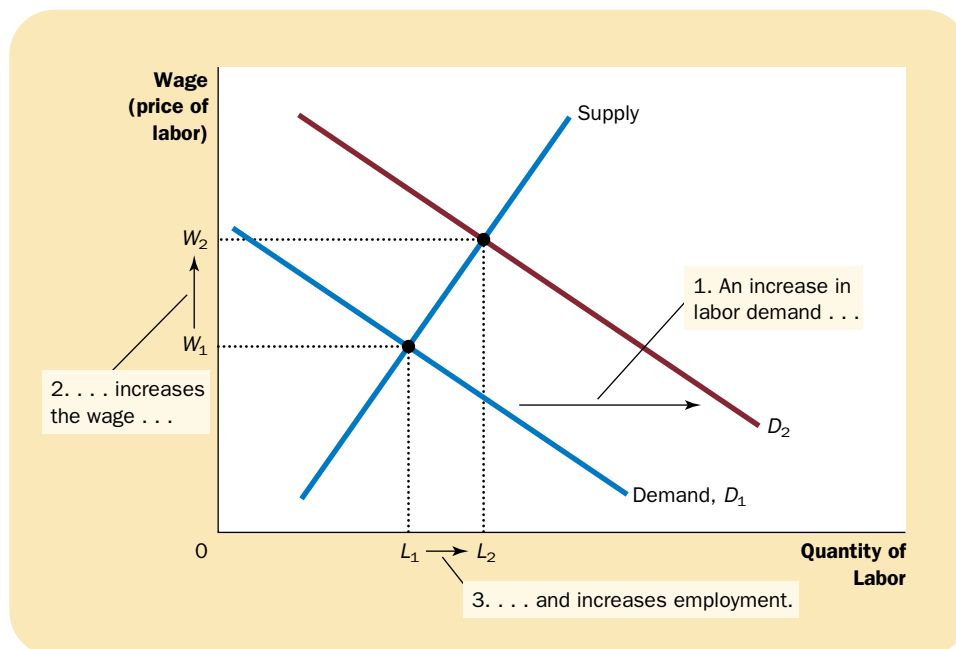
SHIFTS IN LABOR DEMAND

Now suppose that an increase in the popularity of apples causes their price to rise. This price increase does not change the marginal product of labor for any given number of workers, but it does raise the *value* of the marginal product. With a higher price of apples, hiring more apple pickers is now profitable. As Figure 18-6 shows, when the demand for labor shifts to the right from D_1 to D_2 , the equilibrium wage rises from W_1 to W_2 , and equilibrium employment rises from L_1 to L_2 . Once again, the wage and the value of the marginal product of labor move together.

This analysis shows that prosperity for firms in an industry is often linked to prosperity for workers in that industry. When the price of apples rises, apple

Figure 18-6

A SHIFT IN LABOR DEMAND. When labor demand increases from D_1 to D_2 , perhaps because of an increase in the price of the firms' output, the equilibrium wage rises from W_1 to W_2 , and employment rises from L_1 to L_2 . Again, the change in the wage reflects a change in the value of the marginal product of labor: With a higher output price, the added output from an extra worker is more valuable.



producers make greater profit, and apple pickers earn higher wages. When the price of apples falls, apple producers earn smaller profit, and apple pickers earn lower wages. This lesson is well known to workers in industries with highly volatile prices. Workers in oil fields, for instance, know from experience that their earnings are closely linked to the world price of crude oil.

From these examples, you should now have a good understanding of how wages are set in competitive labor markets. Labor supply and labor demand together determine the equilibrium wage, and shifts in the supply or demand curve for labor cause the equilibrium wage to change. At the same time, profit maximization by the firms that demand labor ensures that the equilibrium wage always equals the value of the marginal product of labor.

CASE STUDY PRODUCTIVITY AND WAGES



One of the *Ten Principles of Economics* in Chapter 1 is that our standard of living depends on our ability to produce goods and services. We can now see how this principle works in the market for labor. In particular, our analysis of labor demand shows that wages equal productivity as measured by the value of the marginal product of labor. Put simply, highly productive workers are highly paid, and less productive workers are less highly paid.

This lesson is key to understanding why workers today are better off than workers in previous generations. Table 18-2 presents some data on growth in productivity and growth in wages (adjusted for inflation). From 1959 to 1997, productivity as measured by output per hour of work grew about 1.8 percent per year; at this rate, productivity doubles about every 40 years. Over this period, wages grew at a similar rate of 1.7 percent per year.

Table 18-2

PRODUCTIVITY AND WAGE
GROWTH IN THE UNITED STATES

TIME PERIOD	GROWTH RATE OF PRODUCTIVITY	GROWTH RATE OF REAL WAGES
1959–1997	1.8	1.7
1959–1973	2.9	2.9
1973–1997	1.1	1.0

SOURCE: *Economic Report of the President 1999*, table B-49, p. 384. Growth in productivity is measured here as the annualized rate of change in output per hour in the nonfarm business sector. Growth in real wages is measured as the annualized change in compensation per hour in the nonfarm business sector divided by the implicit price deflator for that sector. These productivity data measure average productivity—the quantity of output divided by the quantity of labor—rather than marginal productivity, but average and marginal productivity are thought to move closely together.

Table 18-2 also shows that, beginning around 1973, growth in productivity slowed from 2.9 to 1.1 percent per year. This 1.8 percentage-point slowdown in productivity coincided with a slowdown in wage growth of 1.9 percentage points. Because of this productivity slowdown, workers in the 1980s and 1990s did not experience the same rapid growth in living standards that their parents enjoyed. A slowdown of 1.8 percentage points might not seem large, but accumulated over many years, even a small change in a growth rate is significant. If productivity and wages had grown at the same rate since 1973 as they did previously, workers' earnings would now be about 50 percent higher than they are.

The link between productivity and wages also sheds light on international experience. Table 18-3 presents some data on productivity growth and wage growth for a representative group of countries, ranked in order of their productivity growth. Although these international data are far from precise, a close link between the two variables is apparent. In South Korea, Hong Kong, and Singapore, productivity has grown rapidly, and so have wages. In Mexico, Argentina, and Iran, productivity has fallen, and so have wages. The United States falls about in the middle of the distribution: By international standards, U.S. productivity growth and wage growth have been neither exceptionally bad nor exceptionally good.

What causes productivity and wages to vary so much over time and across countries? A complete answer to this question requires an analysis of long-run economic growth, a topic beyond the scope of this chapter. We can, however, briefly note three key determinants of productivity:

- ◆ *Physical capital*: When workers work with a larger quantity of equipment and structures, they produce more.
- ◆ *Human capital*: When workers are more educated, they produce more.
- ◆ *Technological knowledge*: When workers have access to more sophisticated technologies, they produce more.

Physical capital, human capital, and technological knowledge are the ultimate sources of most of the differences in productivity, wages, and standards of living.

Table 18-3

PRODUCTIVITY AND WAGE
GROWTH AROUND THE WORLD

COUNTRY	GROWTH RATE OF PRODUCTIVITY	GROWTH RATE OF REAL WAGES
South Korea	8.5	7.9
Hong Kong	5.5	4.9
Singapore	5.3	5.0
Indonesia	4.0	4.4
Japan	3.6	2.0
India	3.1	3.4
United Kingdom	2.4	2.4
United States	1.7	0.5
Brazil	0.4	-2.4
Mexico	-0.2	-3.0
Argentina	-0.9	-1.3
Iran	-1.4	-7.9

SOURCE: *World Development Report 1994*, table 1, pp. 162–163, and table 7, pp. 174–175. Growth in productivity is measured here as the annualized rate of change in gross national product per person from 1980 to 1992. Growth in wages is measured as the annualized change in earnings per employee in manufacturing from 1980 to 1991.

QUICK QUIZ: How does an immigration of workers affect labor supply, labor demand, the marginal product of labor, and the equilibrium wage?

THE OTHER FACTORS OF PRODUCTION: LAND AND CAPITAL

We have seen how firms decide how much labor to hire and how these decisions determine workers' wages. At the same time that firms are hiring workers, they are also deciding about other inputs to production. For example, our apple-producing firm might have to choose the size of its apple orchard and the number of ladders to make available to its apple pickers. We can think of the firm's factors of production as falling into three categories: labor, land, and capital.

capital

the equipment and structures used to produce goods and services

The meaning of the terms *labor* and *land* is clear, but the definition of *capital* is somewhat tricky. Economists use the term **capital** to refer to the stock of equipment and structures used for production. That is, the economy's capital represents the accumulation of goods produced in the past that are being used in the present to produce new goods and services. For our apple firm, the capital stock includes the ladders used to climb the trees, the trucks used to transport the apples, the buildings used to store the apples, and even the trees themselves.

EQUILIBRIUM IN THE MARKETS FOR LAND AND CAPITAL

What determines how much the owners of land and capital earn for their contribution to the production process? Before answering this question, we need to

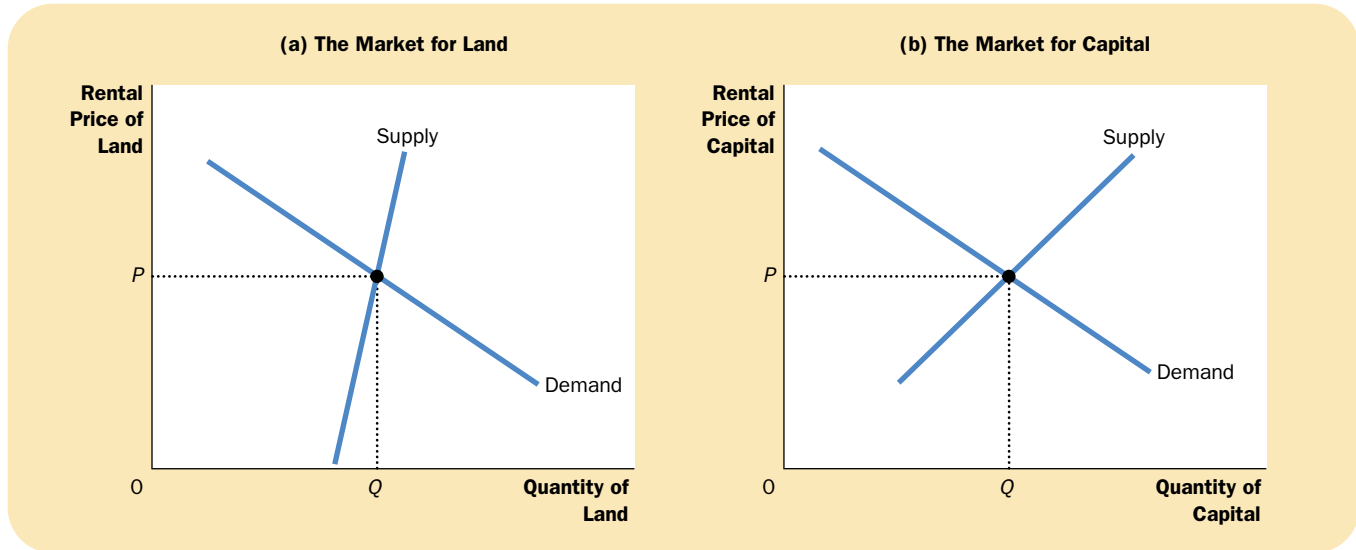


Figure 18-7

THE MARKETS FOR LAND AND CAPITAL. Supply and demand determine the compensation paid to the owners of land, as shown in panel (a), and the compensation paid to the owners of capital, as shown in panel (b). The demand for each factor, in turn, depends on the value of the marginal product of that factor.

distinguish between two prices: the purchase price and the rental price. The *purchase price* of land or capital is the price a person pays to own that factor of production indefinitely. The *rental price* is the price a person pays to use that factor for a limited period of time. It is important to keep this distinction in mind because, as we will see, these prices are determined by somewhat different economic forces.

Having defined these terms, we can now apply the theory of factor demand we developed for the labor market to the markets for land and capital. The wage is, after all, simply the rental price of labor. Therefore, much of what we have learned about wage determination applies also to the rental prices of land and capital. As Figure 18-7 illustrates, the rental price of land, shown in panel (a), and the rental price of capital, shown in panel (b), are determined by supply and demand. Moreover, the demand for land and capital is determined just like the demand for labor. That is, when our apple-producing firm is deciding how much land and how many ladders to rent, it follows the same logic as when deciding how many workers to hire. For both land and capital, the firm increases the quantity hired until the value of the factor's marginal product equals the factor's price. Thus, the demand curve for each factor reflects the marginal productivity of that factor.

We can now explain how much income goes to labor, how much goes to landowners, and how much goes to the owners of capital. As long as the firms using the factors of production are competitive and profit-maximizing, each factor's rental price must equal the value of the marginal product for that factor. *Labor, land, and capital each earn the value of their marginal contribution to the production process.*

Now consider the purchase price of land and capital. The rental price and the purchase price are obviously related: Buyers are willing to pay more to buy a piece of land or capital if it produces a valuable stream of rental income. And, as we

FYI*What Is
Capital Income?*

Labor income is an easy concept to understand: It is the paycheck that workers get from their employers. The income earned by capital, however, is less obvious.

In our analysis, we have been implicitly assuming that households own the economy's stock of capital—ladders, drill presses, warehouses, etc.—and rent it to the firms that use it. Capital income, in this case,

is the rent that households receive for the use of their capital. This assumption simplified our analysis of how capital owners are compensated, but it is not entirely realistic. In fact, firms usually own the capital they use and, therefore, they receive the earnings from this capital.

These earnings from capital, however, eventually get paid to households. Some of the earnings are paid in the form of interest to those households who have lent money to firms. Bondholders and bank depositors are two examples of recipients of interest. Thus, when you receive inter-

est on your bank account, that income is part of the economy's capital income.

In addition, some of the earnings from capital are paid to households in the form of dividends. Dividends are payments by a firm to the firm's stockholders. A stockholder is a person who has bought a share in the ownership of the firm and, therefore, is entitled to share in the firm's profits.

A firm does not have to pay out all of its earnings to households in the form of interest and dividends. Instead, it can retain some earnings within the firm and use these earnings to buy additional capital. Although these retained earnings do not get paid to the firm's stockholders, the stockholders benefit from them nonetheless. Because retained earnings increase the amount of capital the firm owns, they tend to increase future earnings and, thereby, the value of the firm's stock.

These institutional details are interesting and important, but they do not alter our conclusion about the income earned by the owners of capital. Capital is paid according to the value of its marginal product, regardless of whether this income gets transmitted to households in the form of interest or dividends or whether it is kept within firms as retained earnings.

have just seen, the equilibrium rental income at any point in time equals the value of that factor's marginal product. Therefore, the equilibrium purchase price of a piece of land or capital depends on both the current value of the marginal product and the value of the marginal product expected to prevail in the future.

LINKAGES AMONG THE FACTORS OF PRODUCTION

We have seen that the price paid to any factor of production—labor, land, or capital—equals the value of the marginal product of that factor. The marginal product of any factor, in turn, depends on the quantity of that factor that is available. Because of diminishing returns, a factor in abundant supply has a low marginal product and thus a low price, and a factor in scarce supply has a high marginal product and a high price. As a result, when the supply of a factor falls, its equilibrium factor price rises.

When the supply of any factor changes, however, the effects are not limited to the market for that factor. In most situations, factors of production are used together in a way that makes the productivity of each factor dependent on the quantities of the other factors available to be used in the production process. As a result, a change in the supply of any one factor alters the earnings of all the factors.

For example, suppose that a hurricane destroys many of the ladders that workers use to pick apples from the orchards. What happens to the earnings of the various factors of production? Most obviously, the supply of ladders falls and,

therefore, the equilibrium rental price of ladders rises. Those owners who were lucky enough to avoid damage to their ladders now earn a higher return when they rent out their ladders to the firms that produce apples.

Yet the effects of this event do not stop at the ladder market. Because there are fewer ladders with which to work, the workers who pick apples have a smaller marginal product. Thus, the reduction in the supply of ladders reduces the demand for the labor of apple pickers, and this causes the equilibrium wage to fall.

This story shows a general lesson: An event that changes the supply of any factor of production can alter the earnings of all the factors. The change in earnings of any factor can be found by analyzing the impact of the event on the value of the marginal product of that factor.

CASE STUDY THE ECONOMICS OF THE BLACK DEATH

In fourteenth-century Europe, the bubonic plague wiped out about one-third of the population within a few years. This event, called the *Black Death*, provides a grisly natural experiment to test the theory of factor markets that we have just developed. Consider the effects of the Black Death on those who were lucky enough to survive. What do you think happened to the wages earned by workers and the rents earned by landowners?

To answer this question, let's examine the effects of a reduced population on the marginal product of labor and the marginal product of land. With a smaller supply of workers, the marginal product of labor rises. (This is simply diminishing marginal product working in reverse.) Thus, we would expect the Black Death to raise wages.

Because land and labor are used together in production, a smaller supply of workers also affects the market for land, the other major factor of production in medieval Europe. With fewer workers available to farm the land, an additional unit of land produced less additional output. In other words, the marginal product of land fell. Thus, we would expect the Black Death to lower rents.

In fact, both predictions are consistent with the historical evidence. Wages approximately doubled during this period, and rents declined 50 percent or more. The Black Death led to economic prosperity for the peasant classes and reduced incomes for the landed classes.



WORKERS WHO SURVIVED THE PLAGUE WERE LUCKY IN MORE WAYS THAN ONE.

QUICK QUIZ: What determines the income of the owners of land and capital? ♦ How would an increase in the quantity of capital affect the incomes of those who already own capital? How would it affect the incomes of workers?

CONCLUSION

This chapter explained how labor, land, and capital are compensated for the roles they play in the production process. The theory developed here is called the *neoclassical theory of distribution*. According to the neoclassical theory, the amount paid to each factor of production depends on the supply and demand for that factor.

The demand, in turn, depends on that particular factor's marginal productivity. In equilibrium, each factor of production earns the value of its marginal contribution to the production of goods and services.

The neoclassical theory of distribution is widely accepted. Most economists begin with the neoclassical theory when trying to explain how the U.S. economy's \$8 trillion of income is distributed among the economy's various members. In the following two chapters, we consider the distribution of income in more detail. As you will see, the neoclassical theory provides the framework for this discussion.

Even at this point you can use the theory to answer the question that began this chapter: Why are computer programmers paid more than gas station attendants? It is because programmers can produce a good of greater market value than can a gas station attendant. People are willing to pay dearly for a good computer game, but they are willing to pay little to have their gas pumped and their windshield washed. The wages of these workers reflect the market prices of the goods they produce. If people suddenly got tired of using computers and decided to spend more time driving, the prices of these goods would change, and so would the equilibrium wages of these two groups of workers.

Summary

- ◆ The economy's income is distributed in the markets for the factors of production. The three most important factors of production are labor, land, and capital.
- ◆ The demand for factors, such as labor, is a derived demand that comes from firms that use the factors to produce goods and services. Competitive, profit-maximizing firms hire each factor up to the point at which the value of the marginal product of the factor equals its price.
- ◆ The supply of labor arises from individuals' tradeoff between work and leisure. An upward-sloping labor supply curve means that people respond to an increase in the wage by enjoying less leisure and working more hours.
- ◆ The price paid to each factor adjusts to balance the supply and demand for that factor. Because factor demand reflects the value of the marginal product of that factor, in equilibrium each factor is compensated according to its marginal contribution to the production of goods and services.
- ◆ Because factors of production are used together, the marginal product of any one factor depends on the quantities of all factors that are available. As a result, a change in the supply of one factor alters the equilibrium earnings of all the factors.

Key Concepts

factors of production, p. 398
production function, p. 400

marginal product of labor, p. 400
diminishing marginal product, p. 401

value of the marginal product, p. 401
capital, p. 410

Questions for Review

1. Explain how a firm's production function is related to its marginal product of labor, how a firm's marginal product of labor is related to the value of its marginal product, and how a firm's value of marginal product is related to its demand for labor.

2. Give two examples of events that could shift the demand for labor.
3. Give two examples of events that could shift the supply of labor.
4. Explain how the wage can adjust to balance the supply and demand for labor while simultaneously equaling the value of the marginal product of labor.
5. If the population of the United States suddenly grew because of a large immigration, what would happen to wages? What would happen to the rents earned by the owners of land and capital?

Problems and Applications

1. Suppose that the president proposes a new law aimed at reducing health care costs: All Americans are to be required to eat one apple daily.
 - a. How would this apple-a-day law affect the demand and equilibrium price of apples?
 - b. How would the law affect the marginal product and the value of the marginal product of apple pickers?
 - c. How would the law affect the demand and equilibrium wage for apple pickers?
2. Henry Ford once said: "It is not the employer who pays wages—he only handles the money. It is the product that pays wages." Explain.
3. Show the effect of each of the following events on the market for labor in the computer manufacturing industry.
 - a. Congress buys personal computers for all American college students.
 - b. More college students major in engineering and computer science.
 - c. Computer firms build new manufacturing plants.
4. Your enterprising uncle opens a sandwich shop that employs 7 people. The employees are paid \$6 per hour, and a sandwich sells for \$3. If your uncle is maximizing his profit, what is the value of the marginal product of the last worker he hired? What is that worker's marginal product?
5. Imagine a firm that hires two types of workers—some with computer skills and some without. If technology advances, so that computers become more useful to the firm, what happens to the marginal product of the two types? What happens to equilibrium wages? Explain, using appropriate diagrams.
6. Suppose a freeze in Florida destroys part of the Florida orange crop.
 - a. Explain what happens to the price of oranges and the marginal product of orange pickers as a result of the freeze. Can you say what happens to the demand for orange pickers? Why or why not?
- b. Suppose the price of oranges doubles and the marginal product falls by 30 percent. What happens to the equilibrium wage of orange pickers?
- c. Suppose the price of oranges rises by 30 percent and the marginal product falls by 50 percent. What happens to the equilibrium wage of orange pickers?
7. During the 1980s and 1990s the United States experienced a significant inflow of capital from other countries. For example, Toyota, BMW, and other foreign car companies built auto plants in the United States.
 - a. Using a diagram of the U.S. capital market, show the effect of this inflow on the rental price of capital in the United States and on the quantity of capital in use.
 - b. Using a diagram of the U.S. labor market, show the effect of the capital inflow on the average wage paid to U.S. workers.
8. Suppose that labor is the only input used by a perfectly competitive firm that can hire workers for \$50 per day. The firm's production function is as follows:

DAYS OF LABOR	UNITS OF OUTPUT
0	0
1	7
2	13
3	19
4	25
5	28
6	29

Each unit of output sells for \$10. Plot the firm's demand for labor. How many days of labor should the firm hire? Show this point on your graph.

9. (This question is challenging.) In recent years some policymakers have proposed requiring firms to give workers certain fringe benefits. For example, in 1993 President Clinton proposed requiring firms to provide health insurance to their workers. Let's consider the effects of such a policy on the labor market.
 - a. Suppose that a law required firms to give each worker \$3 of fringe benefits for every hour that the worker is employed by the firm. How does this law affect the marginal profit that a firm earns from each worker? How does the law affect the demand curve for labor? Draw your answer on a graph with the cash wage on the vertical axis.
 - b. If there is no change in labor supply, how would this law affect employment and wages?
 - c. Why might the labor supply curve shift in response to this law? Would this shift in labor supply raise or lower the impact of the law on wages and employment?
 - d. As Chapter 6 discussed, the wages of some workers, particularly the unskilled and inexperienced, are kept above the equilibrium level by minimum-wage laws. What effect would a fringe-benefit mandate have for these workers?
10. (This question is challenging.) This chapter has assumed that labor is supplied by individual workers acting competitively. In some markets, however, the supply of labor is determined by a union of workers.
 - a. Explain why the situation faced by a labor union may resemble the situation faced by a monopoly firm.
 - b. The goal of a monopoly firm is to maximize profits. Is there an analogous goal for labor unions?
 - c. Now extend the analogy between monopoly firms and unions. How do you suppose that the wage set by a union compares to the wage in a competitive market? How do you suppose employment differs in the two cases?
 - d. What other goals might unions have that make unions different from monopoly firms?

19



EARNINGS AND DISCRIMINATION

In the United States today, the typical physician earns about \$200,000 a year, the typical police officer about \$50,000, and the typical farmworker about \$20,000. These examples illustrate the large differences in earnings that are so common in our economy. These differences explain why some people live in mansions, ride in limousines, and vacation on the French Riviera, while other people live in small apartments, ride the bus, and vacation in their own back yards.

Why do earnings vary so much from person to person? Chapter 18, which developed the basic neoclassical theory of the labor market, offers an answer to this question. There we saw that wages are governed by labor supply and labor demand. Labor demand, in turn, reflects the marginal productivity of labor. In equilibrium, each worker is paid the value of his or her marginal contribution to the economy's production of goods and services.

IN THIS CHAPTER YOU WILL . . .

Examine how wages compensate for differences in job characteristics

Learn and compare the human-capital and signaling theories of education

Examine why in some occupations a few superstars earn tremendous incomes

Learn why wages rise above the level that balances supply and demand

Consider why it is difficult to measure the impact of discrimination on wages

See when market forces can and cannot provide a natural remedy for discrimination

Consider the debate over comparable worth as a system for setting wages

This theory of the labor market, though widely accepted by economists, is only the beginning of the story. To understand the wide variation in earnings that we observe, we must go beyond this general framework and examine more precisely what determines the supply and demand for different types of labor. That is our goal in this chapter.

SOME DETERMINANTS OF EQUILIBRIUM WAGES

Workers differ from one another in many ways. Jobs also have differing characteristics—both in terms of the wage they pay and in terms of their nonmonetary attributes. In this section we consider how the characteristics of workers and jobs affect labor supply, labor demand, and equilibrium wages.

COMPENSATING DIFFERENTIALS

When a worker is deciding whether to take a job, the wage is only one of many job attributes that the worker takes into account. Some jobs are easy, fun, and safe; others are hard, dull, and dangerous. The better the job as gauged by these nonmonetary characteristics, the more people there are who are willing to do the job at any



"On the one hand, I know I could make more money if I left public service for the private sector, but, on the other hand, I couldn't chop off heads."

given wage. In other words, the supply of labor for easy, fun, and safe jobs is greater than the supply of labor for hard, dull, and dangerous jobs. As a result, “good” jobs will tend to have lower equilibrium wages than “bad” jobs.

For example, imagine you are looking for a summer job in the local beach community. Two kinds of jobs are available. You can take a job as a beach-badge checker, or you can take a job as a garbage collector. The beach-badge checkers take leisurely strolls along the beach during the day and check to make sure the tourists have bought the required beach permits. The garbage collectors wake up before dawn to drive dirty, noisy trucks around town to pick up garbage. Which job would you want? Most people would prefer the beach job if the wages were the same. To induce people to become garbage collectors, the town has to offer higher wages to garbage collectors than to beach-badge checkers.

Economists use the term **compensating differential** to refer to a difference in wages that arises from nonmonetary characteristics of different jobs. Compensating differentials are prevalent in the economy. Here are some examples:

- ◆ Coal miners are paid more than other workers with similar levels of education. Their higher wage compensates them for the dirty and dangerous nature of coal mining, as well as the long-term health problems that coal miners experience.
- ◆ Workers who work the night shift at factories are paid more than similar workers who work the day shift. The higher wage compensates them for having to work at night and sleep during the day, a lifestyle that most people find undesirable.
- ◆ Professors are paid less than lawyers and doctors, who have similar amounts of education. Professors’ lower wages compensate them for the great intellectual and personal satisfaction that their jobs offer. (Indeed, teaching economics is so much fun that it is surprising that economics professors get paid anything at all!)

compensating differential

a difference in wages that arises to offset the nonmonetary characteristics of different jobs

HUMAN CAPITAL

As we discussed in the previous chapter, the word *capital* usually refers to the economy’s stock of equipment and structures. The capital stock includes the farmer’s tractor, the manufacturer’s factory, and the teacher’s blackboard. The essence of capital is that it is a factor of production that itself has been produced.

There is another type of capital that, while less tangible than physical capital, is just as important to the economy’s production. **Human capital** is the accumulation of investments in people. The most important type of human capital is education. Like all forms of capital, education represents an expenditure of resources at one point in time to raise productivity in the future. But, unlike an investment in other forms of capital, an investment in education is tied to a specific person, and this linkage is what makes it human capital.

Not surprisingly, workers with more human capital on average earn more than those with less human capital. College graduates in the United States, for example, earn about twice as much as those workers who end their education with a high school diploma. This large difference has been documented in many countries around the world. It tends to be even larger in less developed countries, where educated workers are in scarce supply.

human capital

the accumulation of investments in people, such as education and on-the-job training

It is easy to see why education raises wages from the perspective of supply and demand. Firms—the demanders of labor—are willing to pay more for the highly educated because highly educated workers have higher marginal products. Workers—the suppliers of labor—are willing to pay the cost of becoming educated only if there is a reward for doing so. In essence, the difference in wages between highly educated workers and less educated workers may be considered a compensating differential for the cost of becoming educated.

CASE STUDY THE INCREASING VALUE OF SKILLS

“The rich get richer, and the poor get poorer.” Like many adages, this one is not always true, but recently it has been. Many studies have documented that the earnings gap between workers with high skills and workers with low skills has increased over the past two decades.

Table 19-1 presents data on the average earnings of college graduates and of high school graduates without any additional education. These data show the increase in the financial reward from education. In 1978, a man on average earned 66 percent more with a college degree than without one; by 1998, this figure had risen to 118 percent. For woman, the reward for attending college rose from a 55 percent increase in earnings to a 98 percent increase. The incentive to stay in school is as great today as it has ever been.

Why has the gap in earnings between skilled and unskilled workers risen in recent years? No one knows for sure, but economists have proposed two hypotheses to explain this trend. Both hypotheses suggest that the demand for skilled labor has risen over time relative to the demand for unskilled labor. The shift in demand has led to a corresponding change in wages, which in turn has led to greater inequality.

The first hypothesis is that international trade has altered the relative demand for skilled and unskilled labor. In recent years, the amount of trade with other countries has increased substantially. Imports into the United States have risen from 5 percent of total U.S. production in 1970 to 13 percent in 1998. Exports from the United States have risen from 6 percent in 1970 to 11 percent in 1998. Because unskilled labor is plentiful and cheap in many foreign countries,

Table 19-1

AVERAGE ANNUAL EARNINGS BY EDUCATIONAL ATTAINMENT. College graduates have always earned more than workers without the benefit of college, but the salary gap grew even larger during the 1980s and 1990s.

		1978	1998
MEN	High school, no college	\$31,847	\$28,742
	College graduates	\$52,761	\$62,588
	Percent extra for college grads	+66%	+118%
WOMEN	High school, no college	\$14,953	\$17,898
	College graduates	\$23,170	\$35,431
	Percent extra for college grads	+55%	+98%

NOTE: Earnings data are adjusted for inflation and are expressed in 1998 dollars. Data apply to workers age 18 and over.

SOURCE: U.S. Census Bureau.

the United States tends to import goods produced with unskilled labor and export goods produced with skilled labor. Thus, when international trade expands, the domestic demand for skilled labor rises, and the domestic demand for unskilled labor falls.

The second hypothesis is that changes in technology have altered the relative demand for skilled and unskilled labor. Consider, for instance, the introduction of computers. Computers raise the demand for skilled workers who can use the new machines and reduce the demand for the unskilled workers whose jobs are replaced by the computers. For example, many companies now rely more on computer databases, and less on filing cabinets, to keep business records. This change raises the demand for computer programmers and reduces the demand for filing clerks. Thus, as more firms begin to use computers, the demand for skilled labor rises, and the demand for unskilled labor falls.

Economists have found it difficult to gauge the validity of these two hypotheses. It is possible, of course, that both are true: Increasing international trade and technological change may share responsibility for the increasing inequality we have observed in recent decades.

ABILITY, EFFORT, AND CHANCE

Why do major league baseball players get paid more than minor league players? Certainly, the higher wage is not a compensating differential. Playing in the major leagues is not a less pleasant task than playing in the minor leagues; in fact, the opposite is true. The major leagues do not require more years of schooling or more experience. To a large extent, players in the major leagues earn more just because they have greater natural ability.

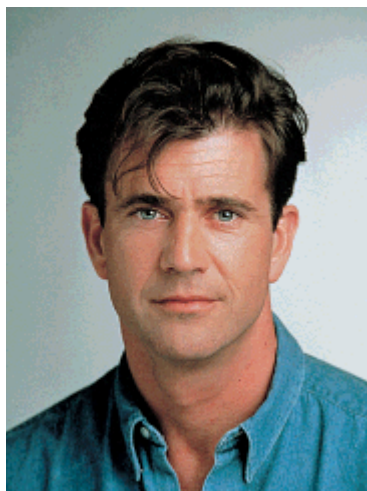
Natural ability is important for workers in all occupations. Because of heredity and upbringing, people differ in their physical and mental attributes. Some people are strong, others weak. Some people are smart, others less so. Some people are outgoing, others awkward in social situations. These and many other personal characteristics determine how productive workers are and, therefore, play a role in determining the wages they earn.

Closely related to ability is effort. Some people work hard; others are lazy. We should not be surprised to find that those who work hard are more productive and earn higher wages. To some extent, firms reward workers directly by paying people on the basis of what they produce. Salespeople, for instance, are often paid as a percentage of the sales they make. At other times, hard work is rewarded less directly in the form of a higher annual salary or a bonus.

Chance also plays a role in determining wages. If a person attended a trade school to learn how to repair televisions with vacuum tubes and then found this skill made obsolete by the invention of solid-state electronics, he or she would end up earning a low wage compared to others with similar years of training. The low wage of this worker is due to chance—a phenomenon that economists recognize but do not shed much light on.

How important are ability, effort, and chance in determining wages? It is hard to say, because ability, effort, and chance are hard to measure. But indirect evidence suggests that they are very important. When labor economists study wages, they relate a worker's wage to those variables that can be measured—years of schooling, years of experience, age, and job characteristics. Although all of these

measured variables affect a worker's wage as theory predicts, they account for less than half of the variation in wages in our economy. Because so much of the variation in wages is left unexplained, omitted variables, including ability, effort, and chance, must play an important role.



GOOD LOOKS PAY.

CASE STUDY THE BENEFITS OF BEAUTY

People differ in many ways. One difference is in how attractive they are. The actor Mel Gibson, for instance, is a handsome man. In part for this reason, his movies attract large audiences. Not surprisingly, the large audiences mean a large income for Mr. Gibson.

How prevalent are the economic benefits of beauty? Labor economists Daniel Hamermesh and Jeff Biddle tried to answer this question in a study published in the December 1994 issue of the *American Economic Review*. Hamermesh and Biddle examined data from surveys of individuals in the United States and Canada. The interviewers who conducted the survey were asked to rate each respondent's physical appearance. Hamermesh and Biddle then examined how much the wages of the respondents depended on the standard determinants—education, experience, and so on—and how much they depended on physical appearance.

Hamermesh and Biddle found that beauty pays. People who are deemed to be more attractive than average earn 5 percent more than people of average looks. People of average looks earn 5 to 10 percent more than people considered less attractive than average. Similar results were found for men and women.

What explains these differences in wages? There are several ways to interpret the “beauty premium.”

One interpretation is that good looks are themselves a type of innate ability determining productivity and wages. Some people are born with the attributes of a movie star; other people are not. Good looks are useful in any job in which workers present themselves to the public—such as acting, sales, and waiting on tables. In this case, an attractive worker is more valuable to the firm than an unattractive worker. The firm's willingness to pay more to attractive workers reflects its customers' preferences.

A second interpretation is that reported beauty is an indirect measure of other types of ability. How attractive a person appears depends on more than just heredity. It also depends on dress, hairstyle, personal demeanor, and other attributes that a person can control. Perhaps a person who successfully projects an attractive image in a survey interview is more likely to be an intelligent person who succeeds at other tasks as well.

A third interpretation is that the beauty premium is a type of discrimination, a topic to which we return later.

AN ALTERNATIVE VIEW OF EDUCATION: SIGNALING

Earlier we discussed the human-capital view of education, according to which schooling raises workers' wages because it makes them more productive. Although this view is widely accepted, some economists have proposed an alternative theory, which emphasizes that firms use educational attainment as a way of sorting between high-ability and low-ability workers. According to this alternative

view, when people earn a college degree, for instance, they do not become more productive, but they do *signal* their high ability to prospective employers. Because it is easier for high-ability people to earn a college degree than it is for low-ability people, more high-ability people get college degrees. As a result, it is rational for firms to interpret a college degree as a signal of ability.

The signaling theory of education is similar to the signaling theory of advertising discussed in Chapter 17. In the signaling theory of advertising, the advertisement itself contains no real information, but the firm signals the quality of its product to consumers by its willingness to spend money on advertising. In the signaling theory of education, schooling has no real productivity benefit, but the worker signals his innate productivity to employers by his willingness to spend years at school. In both cases, an action is being taken not for its intrinsic benefit but because the willingness to take that action conveys private information to someone observing it.

Thus, we now have two views of education: the human-capital theory and the signaling theory. Both views can explain why more educated workers tend to earn more than less educated workers. According to the human-capital view, education makes workers more productive; according to the signaling view, education is correlated with natural ability. But the two views have radically different predictions for the effects of policies that aim to increase educational attainment. According to the human-capital view, increasing educational levels for all workers would raise all workers' productivity and thereby their wages. According to the signaling view, education does not enhance productivity, so raising all workers' educational levels would not affect wages.

Most likely, truth lies somewhere between these two extremes. The benefits to education are probably a combination of the productivity-enhancing effects of human capital and the productivity-revealing effects of signaling. The open question is the relative size of these two effects.

CASE STUDY HUMAN CAPITAL, NATURAL ABILITY, AND COMPULSORY SCHOOL ATTENDANCE

Does attending school increase wages because it increases productivity, or does it only appear to increase productivity because high-ability people are more likely to stay in school? This question is important both for judging the various theories of education and for evaluating alternative education policies.

If economists could conduct controlled experiments like laboratory scientists, it would be easy to answer this question. We could choose some experimental subjects from the school-age population and then randomly divide them into various groups. For each group we could require a different amount of school attendance. By comparing the difference in the educational attainment and the difference in subsequent wages of the various groups, we could see whether education does in fact increase productivity. Because the groups would be chosen randomly, we could be sure that the difference in wages was not attributable to a difference in natural ability.

Although conducting such an experiment might seem difficult, the laws of the United States inadvertently provide a natural experiment that is quite similar. All students in the United States are required by law to attend school, but the laws vary from state to state. Some states allow students to drop out at age

16, while others require attendance until age 17 or 18. Moreover, the laws have changed over time. Between 1970 and 1980, for instance, Wyoming reduced the school-attendance age from 17 to 16, while Washington raised it from 16 to 18. This variation across states and over time provides data with which to study the effects of compulsory school attendance.

Even within a state, school-attendance laws have different effects on different people. Students start attending school at different ages, depending on the month of the year in which they were born. Yet all students can drop out as soon as they reach the minimum legal age; they are not required to finish out the school year. As a result, those who start school at a relatively young age are required to spend more time in school than those who start school at a relatively old age. This variation across students within a state also provides a way to study the effects of compulsory attendance.

In an article published in the November 1991 issue of the *Quarterly Journal of Economics*, labor economists Joshua Angrist and Alan Krueger used this natural experiment to study the relationship between schooling and wages. Because the duration of each student's compulsory schooling depends on his or her state of residence and month of birth, and not on natural ability, it was possible to isolate the productivity-enhancing effect of education from the ability-signaling effect. According to Angrist and Krueger's research, those students who were required to finish more school did earn significantly higher subsequent wages than those with lower requirements. This finding indicates that education does raise a worker's productivity, as the human-capital theory suggests.

Although establishing the benefits of compulsory schooling is useful, it does not by itself tell us whether these laws are desirable. That policy judgment requires a more complete analysis of the costs and benefits. At the very least, we would need to compare the benefits of schooling to the opportunity cost—the wages that the student could have earned by dropping out. In addition, requiring a student to stay in school may have external effects on others in society. On the one hand, compulsory school attendance may reduce crime rates, for young dropouts are at high risk of engaging in criminal activity. On the other hand, students who stay in school only because they are required to do so may interfere with the learning of other students who are more committed to their educations.

THE SUPERSTAR PHENOMENON

Although most actors earn very little and often have to take jobs as waiters to support themselves, actor Robin Williams earned \$23 million in 1997. Similarly, although most people who play football do it for free as a hobby, Brett Favre earned \$6.75 million as a pro quarterback. Robin Williams and Brett Favre are superstars in their fields, and their great public appeal is reflected in astronomical incomes.

Why do Robin Williams and Brett Favre earn so much? It is not surprising that there are differences in incomes within occupations. Good carpenters earn more than mediocre carpenters, and good plumbers earn more than mediocre plumbers. People vary in ability and effort, and these differences lead to differences in income. Yet the best carpenters and plumbers do not earn the many millions that are common among the best actors and athletes. What explains the difference?

To understand the tremendous incomes of Robin Williams and Brett Favre, we must examine the special features of the markets in which they sell their services. Superstars arise in markets that have two characteristics:

- ◆ Every customer in the market wants to enjoy the good supplied by the best producer.
- ◆ The good is produced with a technology that makes it possible for the best producer to supply every customer at low cost.

If Robin Williams is the funniest actor around, then everyone will want to see his next movie; seeing twice as many movies by an actor half as funny is not a good substitute. Moreover, it is *possible* for everyone to enjoy the comedy of Robin Williams. Because it is easy to make multiple copies of a film, Robin Williams can provide his service to millions of people simultaneously. Similarly, because football games are broadcast on television, millions of fans can enjoy the extraordinary athletic skills of Brett Favre.

We can now see why there are no superstar carpenters and plumbers. Other things equal, everyone prefers to employ the best carpenter, but a carpenter, unlike a movie actor, can provide his services to only a limited number of customers. Although the best carpenter will be able to command a somewhat higher wage than the average carpenter, the average carpenter will still be able to earn a good living.

ABOVE-EQUILIBRIUM WAGES: MINIMUM-WAGE LAWS, UNIONS, AND EFFICIENCY WAGES

Most analyses of wage differences among workers are based on the equilibrium model of the labor market—that is, wages are assumed to adjust to balance labor supply and labor demand. But this assumption does not always apply. For some workers, wages are set above the level that brings supply and demand into equilibrium. Let's consider three reasons why this might be so.

One reason for above-equilibrium wages is minimum-wage laws, as we first saw in Chapter 6. Most workers in the economy are not affected by these laws because their equilibrium wages are well above the legal minimum. But for some workers, especially the least skilled and experienced, minimum-wage laws raise wages above the level they would earn in an unregulated labor market.

A second reason that wages might rise above their equilibrium level is the market power of labor unions. A **union** is a worker association that bargains with employers over wages and working conditions. Unions often raise wages above the level that would prevail without a union, perhaps because they can threaten to withhold labor from the firm by calling a **strike**. Studies suggest that union workers earn about 10 to 20 percent more than similar nonunion workers.

A third reason for above-equilibrium wages is suggested by the theory of **efficiency wages**. This theory holds that a firm can find it profitable to pay high wages because doing so increases the productivity of its workers. In particular, high wages may reduce worker turnover, increase worker effort, and raise the quality of workers who apply for jobs at the firm. If this theory is correct, then some firms may choose to pay their workers more than they would normally earn.

Above-equilibrium wages, whether caused by minimum-wage laws, unions, or efficiency wages, have similar effects on the labor market. In particular, pushing a wage above the equilibrium level raises the quantity of labor supplied and reduces the quantity of labor demanded. The result is a surplus of labor, or unemployment. The study of unemployment and the public policies aimed to deal with it is usually considered a topic within macroeconomics, so it goes beyond the scope of this chapter. But it would be a mistake to ignore these issues completely

union

a worker association that bargains with employers over wages and working conditions

strike

the organized withdrawal of labor from a firm by a union

efficiency wages

above-equilibrium wages paid by firms in order to increase worker productivity

when analyzing earnings. Although most wage differences can be understood while maintaining the assumption of equilibrium in the labor market, above-equilibrium wages play a role in some cases.

QUICK QUIZ: Define *compensating differential* and give an example.

- ◆ Give two reasons why more educated workers earn more than less educated workers.

THE ECONOMICS OF DISCRIMINATION

discrimination

the offering of different opportunities to similar individuals who differ only by race, ethnic group, sex, age, or other personal characteristics

Another source of differences in wages is discrimination. **Discrimination** occurs when the marketplace offers different opportunities to similar individuals who differ only by race, ethnic group, sex, age, or other personal characteristics. Discrimination reflects some people's prejudice against certain groups in society. Although discrimination is an emotionally charged topic that often generates heated debate, economists try to study the topic objectively in order to separate myth from reality.

MEASURING LABOR-MARKET DISCRIMINATION

How much does discrimination in labor markets affect the earnings of different groups of workers? This question is important, but answering it is not easy.

It might seem natural to gauge the amount of discrimination in labor markets by looking at the average wages of different groups. For instance, in recent years the wage of the average black worker in the United States has been about 20 percent less than the wage of the average white worker. The wage of the average female worker has been about 30 percent less than the wage of the average male worker. These wage differentials are sometimes presented in political debate as evidence that many employers discriminate against blacks and women.

Yet there is an obvious problem with this approach. Even in a labor market free of discrimination, different people have different wages. People differ in the amount of human capital they have and in the kinds of work they are able and willing to do. The wage differences we observe in the economy are, to a large extent, attributable to the determinants of equilibrium wages we discussed in the preceding section. Simply observing differences in wages among broad groups—whites and blacks, men and women—says little about the prevalence of discrimination.

Consider, for example, the role of human capital. About 80 percent of white male workers have a high school diploma, and 25 percent have a college degree. By contrast, only 67 percent of black male workers have a high school diploma, and only 12 percent have a college degree. Thus, at least some of the difference between the wages of whites and the wages of blacks can be traced to differences in educational attainment. Similarly, among white workers, 25 percent of men have a college degree, whereas only 19 percent of women have a college degree, indicating that some of the difference between the wages of men and women is attributable to educational attainment.

In fact, human capital is probably even more important in explaining wage differentials than the foregoing numbers suggest. For many years, public schools in predominantly black areas have been of lower quality—as measured by expenditure, class size, and so on—than public schools in predominantly white areas. Similarly, for many years, schools directed girls away from science and math courses, even though these subjects may have had greater value in the marketplace than some of the alternatives. If we could measure the quality as well as the quantity of education, the differences in human capital among these groups would seem even larger.

Human capital acquired in the form of job experience can also help explain wage differences. In particular, women tend to have less job experience on average than men. One reason is that female labor-force participation has increased over the past several decades. Because of this historic change, the average female worker today is younger than the average male worker. In addition, women are more likely to interrupt their careers to raise children. For both reasons, the experience of the average female worker is less than the experience of the average male worker.

Yet another source of wage differences is compensating differentials. Some analysts have suggested that women take more pleasant jobs on average than men and that this fact explains some of the earnings differential between men and women. For example, women are more likely to be secretaries, and men are more likely to be truck drivers. The relative wages of secretaries and truck drivers depend in part on the working conditions of each job. Because these nonmonetary aspects are hard to measure, it is difficult to gauge the practical importance of compensating differentials in explaining the wage differences that we observe.

In the end, the study of wage differences among groups does not establish any clear conclusion about the prevalence of discrimination in U.S. labor markets. Most economists believe that some of the observed wage differentials are attributable to discrimination, but there is no consensus about how much. The only conclusion about which economists are in consensus is a negative one: *Because the differences in average wages among groups in part reflect differences in human capital and job characteristics, they do not by themselves say anything about how much discrimination there is in the labor market.*

Of course, differences in human capital among groups of workers may themselves reflect discrimination. The inferior schools historically available to black students, for instance, may be traced to prejudice on the part of city councils and school boards. But this kind of discrimination occurs long before the worker enters the labor market. In this case, the disease is political, even if the symptom is economic.

DISCRIMINATION BY EMPLOYERS

Let's now turn from measurement to the economic forces that lie behind discrimination in labor markets. If one group in society receives a lower wage than another group, even after controlling for human capital and job characteristics, who is to blame for this differential?

The answer is not obvious. It might seem natural to blame employers for discriminatory wage differences. After all, employers make the hiring decisions that determine labor demand and wages. If some groups of workers earn lower wages than they should, then it seems that employers are responsible. Yet many economists are

skeptical of this easy answer. They believe that competitive, market economies provide a natural antidote to employer discrimination. That antidote is called the profit motive.

Imagine an economy in which workers are differentiated by their hair color. Blondes and brunettes have the same skills, experience, and work ethic. Yet, because of discrimination, employers prefer not to hire workers with blonde hair. Thus, the demand for blondes is lower than it otherwise would be. As a result, blondes earn a lower wage than brunettes.

How long can this wage differential persist? In this economy, there is an easy way for a firm to beat out its competitors: It can hire blonde workers. By hiring blondes, a firm pays lower wages and thus has lower costs than firms that hire brunettes. Over time, more and more “blonde” firms enter the market to take advantage of this cost advantage. The existing “brunette” firms have higher costs and, therefore, begin to lose money when faced with the new competitors. These losses induce the brunette firms to go out of business. Eventually, the entry of blonde firms and the exit of brunette firms cause the demand for blonde workers to rise and the demand for brunette workers to fall. This process continues until the wage differential disappears.

Put simply, business owners who care only about making money are at an advantage when competing against those who also care about discriminating. As a result, firms that do not discriminate tend to replace those that do. In this way, competitive markets have a natural remedy for employer discrimination.

CASE STUDY **SEGREGATED STREETCARS AND THE PROFIT MOTIVE**

In the early twentieth century, streetcars in many southern cities were segregated by race. White passengers sat in the front of the streetcars, and black passengers sat in the back. What do you suppose caused and maintained this discriminatory practice? And how was this practice viewed by the firms that ran the streetcars?

In a 1986 article in the *Journal of Economic History*, economic historian Jennifer Roback looked at these questions. Roback found that the segregation of races on streetcars was the result of laws that required such segregation. Before these laws were passed, racial discrimination in seating was rare. It was far more common to segregate smokers and nonsmokers.

Moreover, the firms that ran the streetcars often opposed the laws requiring racial segregation. Providing separate seating for different races raised the firms’ costs and reduced their profit. One railroad company manager complained to the city council that, under the segregation laws, “the company has to haul around a good deal of empty space.”

Here is how Roback describes the situation in one southern city:

The railroad company did not initiate the segregation policy and was not at all eager to abide by it. State legislation, public agitation, and a threat to arrest the president of the railroad were all required to induce them to separate the races on their cars. . . . There is no indication that the management was motivated by belief in civil rights or racial equality. The evidence indicates their primary motives were economic; separation was

costly. . . . Officials of the company may or may not have disliked blacks, but they were not willing to forgo the profits necessary to indulge such prejudice.

The story of southern streetcars illustrates a general lesson: Business owners are usually more interested in making profit than in discriminating against a particular group. When firms engage in discriminatory practices, the ultimate source of the discrimination often lies not with the firms themselves but elsewhere. In this particular case, the streetcar companies segregated whites and blacks because discriminatory laws, which the companies opposed, required them to do so.

DISCRIMINATION BY CUSTOMERS AND GOVERNMENTS

Although the profit motive is a strong force acting to eliminate discriminatory wage differentials, there are limits to its corrective abilities. Here we consider two of the most important limits: *customer preferences* and *government policies*.

To see how customer preferences for discrimination can affect wages, consider again our imaginary economy with blondes and brunettes. Suppose that restaurant owners discriminate against blondes when hiring waiters. As a result, blonde waiters earn lower wages than brunette waiters. In this case, a restaurant could open up with blonde waiters and charge lower prices. If customers only cared about the quality and price of their meals, the discriminatory firms would be driven out of business, and the wage differential would disappear.

On the other hand, it is possible that customers prefer being served by brunette waiters. If this preference for discrimination is strong, the entry of blonde restaurants need not succeed in eliminating the wage differential between brunettes and blondes. That is, if customers have discriminatory preferences, a competitive market is consistent with a discriminatory wage differential. An economy with such discrimination would contain two types of restaurants. Blonde restaurants hire blondes, have lower costs, and charge lower prices. Brunette restaurants hire brunettes, have higher costs, and charge higher prices. Customers who did not care about the hair color of their waiters would be attracted to the lower prices at the blonde restaurants. Bigoted customers would go to the brunette restaurants. They would pay for their discriminatory preference in the form of higher prices.

Another way for discrimination to persist in competitive markets is for the government to mandate discriminatory practices. If, for instance, the government passed a law stating that blondes could wash dishes in restaurants but could not work as waiters, then a wage differential could persist in a competitive market. The example of segregated streetcars in the foregoing case study is one example of government-mandated discrimination. More recently, before South Africa abandoned its system of apartheid, blacks were prohibited from working in some jobs. Discriminatory governments pass such laws to suppress the normal equalizing force of free and competitive markets.

To sum up: *Competitive markets contain a natural remedy for employer discrimination. The entry of firms that care only about profit tends to eliminate discriminatory wage differentials. These wage differentials persist in competitive markets only when customers are willing to pay to maintain the discriminatory practice or when the government mandates it.*

IN THE NEWS*Men, Women, and Wages*

WHY DOES THE AVERAGE FEMALE WORKER earn less than the average male worker? In the following article, economist June O'Neill offers some answers to this question.

The Shrinking Pay Gap

BY JUNE ELLENOFF O'NEILL

"Fifty-nine cents," the popular button said, a symbol of the stubborn fact that throughout the post-World War II period, women's wages hovered at around 60 percent of men's, despite an increasing proportion of women working outside the home. This gender gap did not decline through the 1960s and the 1970s

despite the rise of the feminist movement, equal pay and employment legislation, and affirmative action.

But starting in the Reagan years, the gender gap in wages began to decline dramatically. By some measures the ratio of women's earnings to men's rose to nearly 80 percent; and even this number, I believe, overstates the gender gap between men and women with similar skills and training. Why did this dramatic narrowing in relative wages happen?

The answer has less to do with politics or protests than with the realities of the labor market. Although basic skills are acquired in school, it is in the labor market where specialized skills are developed that bring higher wages. During the three decades following World War II women entered the labor market in record numbers. But many of the new entrants had been out of the labor force for considerable periods of time, raising their children. These women diluted the skill level of the rapidly expanding group of employed women. This was the main reason why the gender gap in pay did not narrow during the postwar years.

Today's working women, particularly those younger than forty, are much more nearly equal to men in work experience than were their mothers. Through delayed marriage, low fertility, and an increasing tendency for mothers of young children to work, women have acquired many more years of continuous work experience than was true in the past. (Close to 60 percent of married women with children under age six are now in the labor force; in 1960, the proportion was only 19 percent.)

And the work experience gained by these younger women is likely to have an even greater impact on their future earnings because their work experience has been more correctly anticipated. Many investment choices affecting careers are made at younger ages: years of schooling, subjects in school, other professional training. In the past, women were much less likely than men to invest in lengthy training because they assumed they would not be working enough years to justify it.

In fact, the National Longitudinal Surveys found that even in the late 1960s less than 30 percent of young

CASE STUDY DISCRIMINATION IN SPORTS

As we have seen, measuring discrimination is often difficult. To determine whether one group of workers is discriminated against, a researcher must correct for differences in the productivity between that group and other workers in the economy. Yet, in most firms, it is difficult to measure a particular worker's contribution to the production of goods and services.

One type of firm in which such corrections are easier is the sports team. Professional teams have many objective measures of productivity. In baseball, for instance, we can measure a player's batting average, the frequency of home runs, the number of stolen bases, and so on.

Studies of sports teams suggest that racial discrimination is, in fact, common and that much of the blame lies with customers. One study, published in the *Journal of Labor Economics* in 1988, examined the salaries of basketball players. It found that black players earned 20 percent less than white players of comparable ability. The study also found that attendance at basketball games was larger for teams

women anticipated that they would be working at age thirty-five, yet when this group actually reached thirty-five, more than 70 percent of them were in the labor force. Their underestimation of future work activity surely influenced their early career preparations (or lack thereof). More recent survey data show a dramatic change in expectations. The vast majority of young women now report an intention to work at age thirty-five.

Those changing work expectations are reflected in rising female enrollments in higher education. In 1960, women received 35 percent of all bachelor's degrees in the U.S.; by the 1980s, they received somewhat more than half of them. In 1968, women received 8 percent of the medical degrees, 3 percent of the MBAs, and 4 percent of the law degrees granted that year. In 1986, they received 31 percent of the medical degrees and MBAs and 39 percent of the law degrees. This recent trend in schooling is likely to reinforce the rise in work experience and contribute to continuing increases in the relative earnings of women workers. . . .

Despite the advances of the past decade, women still earn less than men. The hourly earnings of women were 74 percent of the earnings of men in 1992 when ages twenty-five to sixty-four are considered, up from 62 percent in 1979. At ages twenty-five to thirty-four, where women's skills have increased the most, the ratio is 87 percent.

Economist Barbara Bergmann and others attribute the pay gap to "widespread, severe, ongoing discrimination by employers and fellow workers." But discrimination cannot be directly measured. Instead, researchers estimate the extent to which differences in productivity appear to explain the gap and then attribute the rest to discrimination. Such a conclusion is premature, however, when productivity differences are not accurately measured, which is usually the case.

For example, data are seldom available on lifetime patterns of work experience, and even less material is available on factors bearing on work expectations and the intensity and nature of work investments. As these are still the key sources of skill differences between men

and women, there is considerable room for interpretation and disagreement.

When earnings comparisons are restricted to men and women more similar in their experience and life situations, the measured earnings differentials are typically quite small. For example, among people twenty-seven to thirty-three who have never had a child, the earnings of women in the National Longitudinal Survey of Youth are close to 98 percent of men's. . . .

It is true that women and men still do not have the same earnings. But I believe that the differential is largely due to continuing gender differences in the priority placed on market work vs. family responsibilities. Until family roles are more equal, women are not likely to have the same pattern of market work and earnings as men. Technology has reduced the burden of housework, but child care remains a responsibility that is harder to shift to the market.

SOURCE: *The Wall Street Journal*, October 7, 1994, p. A10.

with a greater proportion of white players. One interpretation of these facts is that customer discrimination makes black players less profitable than white players for team owners. In the presence of such customer discrimination, a discriminatory wage gap can persist, even if team owners care only about profit.

A similar situation once existed for baseball players. A study using data from the late 1960s showed that black players earned less than comparable white players. Moreover, fewer fans attended games pitched by blacks than games pitched by whites, even though black pitchers had better records than white pitchers. Studies of more recent salaries in baseball, however, have found no evidence of discriminatory wage differentials.

Another study, published in the *Quarterly Journal of Economics* in 1990, examined the market prices of old baseball cards. This study found similar evidence of discrimination. The cards of black hitters sold for 10 percent less than the cards of comparable white hitters. The cards of black pitchers sold for 13 percent less than the cards of comparable white pitchers. These results suggest customer discrimination among baseball fans.

comparable worth

a doctrine according to which jobs deemed comparable should be paid the same wage

THE DEBATE OVER COMPARABLE WORTH

Should engineers get paid more than librarians? This question is at the heart of the debate over **comparable worth**, a doctrine whereby jobs deemed comparable should be paid the same wage.

Advocates of comparable worth point out that traditionally male occupations have higher wages than traditionally female occupations. They believe that these occupational differences are discriminatory against women. Even if women were paid the same as men for the same type of work, the gender gap in wages would persist until comparable occupations were paid similar wages. Comparable-worth advocates want jobs rated according to a set of impartial criteria—education, experience, responsibility, working conditions, and so on. Under this system, comparably rated jobs would pay the same wage. A librarian with a master’s degree, ten years of experience, and a 40-hour workweek, for instance, would be paid the same as an engineer with a master’s degree, ten years of experience, and a 40-hour workweek.

Most economists are critical of comparable-worth proposals. They argue that a competitive market is the best mechanism for setting wages. It would be nearly impossible, they claim, to measure all of the factors that are relevant for determining the right wage for any job. Moreover, the fact that traditionally female occupations pay less than traditionally male occupations is not by itself evidence of discrimination. Women have in the past spent more time than men raising children. Women are, therefore, more likely to choose occupations that offer flexible hours and other working conditions compatible with child-rearing. To some extent, the gender gap in wages is a compensating differential.

Economists also point out that comparable-worth proposals would have an important unintended side effect. Comparable-worth advocates want the wages in traditionally female occupations to be raised by legal decree. Such a policy would have many of the effects of a minimum wage, which we first discussed in Chapter 6. In particular, when the wage is forced to rise above the equilibrium level, the quantity of labor supplied to these occupations would rise, and the quantity demanded would fall. The result would be higher unemployment in traditionally female occupations. In this way, a comparable-worth law could adversely affect some members of groups that the policy is aimed at helping.

QUICK QUIZ: Why is it hard to establish whether a group of workers is being discriminated against? ♦ Explain how profit-maximizing firms tend to eliminate discriminatory wage differentials. ♦ How might a discriminatory wage differential persist?

CONCLUSION

In competitive markets, workers earn a wage equal to the value of their marginal contribution to the production of goods and services. There are, however, many things that affect the value of the marginal product. Firms pay more for workers who are more talented, more diligent, more experienced, and more educated because these workers are more productive. Firms pay less to those workers against whom customers discriminate because these workers contribute less to revenue.

IN THE NEWS*The Recent Push for
Comparable Worth*

OVER THE PAST SEVERAL YEARS, THE IDEA OF comparable worth—sometimes called pay equity—has made a comeback among some political leaders.

Labor and Women Push for Equal Pay for Equivalent Work

BY MARY LEONARD

WASHINGTON—Nobody says men and women shouldn't get equal pay for doing the same job. But what's brewing now is a big push nationally by the president, organized labor, and women's rights groups to level the gender playing field on wages for different but equivalent work.

In a strategy session today, Senator Tom Harkin, an Iowa Democrat who has been the lonely champion for pay-equity legislation since 1996, will meet with John Podesta, the president's chief of staff, and other officials on ways the White House can boost his bill this year. Yesterday, the AFL-CIO launched a nationwide campaign to pass comparable-worth bills in 24 states, including Massachusetts.

What's going on here? Trying to determine comparable salaries for jobs traditionally held by men and women is an old idea, discredited by some economists as unwieldy, if not downright dumb. They wonder who can and will decide the economic value of a riveter versus a nurse, the comparable pay for a probation officer and a librarian, the equivalent pay for an auto mechanic and a secretary.

Many see pay equity, even if it is difficult to enforce, as the only remedy for wage discrimination, a problem that persists for women, even as they have earned advanced degrees, climbed the corporate ladder, and plopped their children in day care while pursuing full-time jobs in large numbers. . . .

"For too long, working women have been seething while politicians have remained silent," said Karen Nussbaum, director of the AFL-CIO's Working Women's department. "Pay equity can right a long-standing wrong." . . .

Diana Furchtgott-Roth, a resident fellow at the American Enterprise Institute, says there are plenty of reasons why men and women earn different wages—seniority, job risk, and market demand for certain skills—that have nothing to do with discrimination and would not be erased by "cockeyed" pay-equity laws. She said when you compare men and women with the same qualifications doing the same jobs, women earn 95 percent of men's salaries.

"Comparable worth is certainly making a comeback," Furchtgott-Roth said, "and I believe it's because feminists who supported Clinton through the Lewinsky mess are demanding a political payoff. I don't see any Republican support for these proposals."

SOURCE: *The Boston Globe*, February 25, 1999, pp. A1, A22.

The theory of the labor market we have developed in the last two chapters explains why some workers earn higher wages than other workers. The theory does not say that the resulting distribution of income is equal, fair, or desirable in any way. That is the topic we take up in Chapter 20.

Summary

- ◆ Workers earn different wages for many reasons. To some extent, wage differentials compensate workers for job attributes. Other things equal, workers in hard, unpleasant jobs get paid more than workers in easy, pleasant jobs.
- ◆ Workers with more human capital get paid more than workers with less human capital. The return to accumulating human capital is high and has increased over the past decade.
- ◆ Although years of education, experience, and job characteristics affect earnings as theory predicts, there is much variation in earnings that cannot be explained by things that economists can measure. The unexplained variation in earnings is

largely attributable to natural ability, effort, and chance.

- ◆ Some economists have suggested that more educated workers earn higher wages not because education raises productivity but because workers with high natural ability use education as a way to signal their high ability to employers. If this signaling theory were correct, then increasing the educational attainment of all workers would not raise the overall level of wages.
- ◆ Wages are sometimes pushed above the level that brings supply and demand into balance. Three reasons for above-equilibrium wages are minimum-wage laws, unions, and efficiency wages.
- ◆ Some differences in earnings are attributable to discrimination on the basis of race, sex, or other factors.

Measuring the amount of discrimination is difficult, however, because one must correct for differences in human capital and job characteristics.

- ◆ Competitive markets tend to limit the impact of discrimination on wages. If the wages of a group of workers are lower than those of another group for reasons not related to marginal productivity, then nondiscriminatory firms will be more profitable than discriminatory firms. Profit-maximizing behavior, therefore, can act to reduce discriminatory wage differentials. Discrimination can persist in competitive markets if customers are willing to pay more to discriminatory firms or if the government passes laws requiring firms to discriminate.

Key Concepts

compensating differential, p. 419
human capital, p. 419
union, p. 425

strike, p. 425
efficiency wages, p. 425
discrimination, p. 426

comparable worth, p. 432

Questions for Review

1. Why do coal miners get paid more than other workers with similar amounts of education?
2. In what sense is education a type of capital?
3. How might education raise a worker's wage without raising the worker's productivity?
4. What conditions lead to economic superstars? Would you expect to see superstars in dentistry? In music? Explain.
5. Give three reasons why a worker's wage might be above the level that balances supply and demand.
6. What difficulties arise in deciding whether a group of workers has a lower wage because of discrimination?
7. Do the forces of economic competition tend to exacerbate or ameliorate discrimination on the basis of race?
8. Give an example of how discrimination might persist in a competitive market.

Problems and Applications

1. College students sometimes work as summer interns for private firms or the government. Many of these positions pay little or nothing.
 - a. What is the opportunity cost of taking such a job?
 - b. Explain why students are willing to take these jobs.
 - c. If you were to compare the earnings later in life of workers who had worked as interns and those who had taken summer jobs that paid more, what would you expect to find?
2. As explained in Chapter 6, a minimum-wage law distorts the market for low-wage labor. To reduce this distortion, some economists advocate a two-tiered minimum-wage system, with a regular minimum wage for adult workers and a lower, "sub-minimum" wage

for teenage workers. Give two reasons why a single minimum wage might distort the labor market for teenage workers more than it would the market for adult workers.

3. A basic finding of labor economics is that workers who have more experience in the labor force are paid more than workers who have less experience (holding constant the amount of formal education). Why might this be so? Some studies have also found that experience at the same job (called “job tenure”) has an extra positive influence on wages. Explain.
4. At some colleges and universities, economics professors receive higher salaries than professors in some other fields.
 - a. Why might this be true?
 - b. Some other colleges and universities have a policy of paying equal salaries to professors in all fields. At some of these schools, economics professors have lighter teaching loads than professors in some other fields. What role do the differences in teaching loads play?
5. Sara works for Steve, whom she hates because of his snobbish attitude. Yet when she looks for other jobs, the best she can do is find a job paying \$10,000 less than her current salary. Should she take the job? Analyze Sara’s situation from an economic point of view.
6. Imagine that someone were to offer you a choice: You could spend four years studying at the world’s best university, but you would have to keep your attendance there a secret. Or you could be awarded an official degree from the world’s best university, but you couldn’t actually attend. Which choice do you think would enhance your future earnings more? What does your answer say about the debate over signaling versus human capital in the role of education?
7. When recording devices were first invented almost 100 years ago, musicians could suddenly supply their music to large audiences at low cost. How do you suppose this event affected the income of the best musicians? How do you suppose it affected the income of average musicians?
8. Alan runs an economic consulting firm. He hires primarily female economists because, he says, “they will work for less than comparable men because women have fewer job options.” Is Alan’s behavior admirable or despicable? If more employers were like Alan, what would happen to the wage differential between men and women?
9. A case study in this chapter described how customer discrimination in sports seems to have an important effect on players’ earnings. Note that this is possible because sports fans know the players’ characteristics, including their race. Why is this knowledge important for the existence of discrimination? Give some specific examples of industries where customer discrimination is and is not likely to influence wages.
10. Suppose that all young women were channeled into careers as secretaries, nurses, and teachers; at the same time, young men were encouraged to consider these three careers and many others as well.
 - a. Draw a diagram showing the combined labor market for secretaries, nurses, and teachers. Draw a diagram showing the combined labor market for all other fields. In which market is the wage higher? Do men or women receive higher wages on average?
 - b. Now suppose that society changed and encouraged both young women and young men to consider a wide range of careers. Over time, what effect would this change have on the wages in the two markets you illustrated in part (a)? What effect would the change have on the average wages of men and women?
11. Economist June O’Neill argues that “until family roles are more equal, women are not likely to have the same pattern of market work and earnings as men.” What does she mean by the “pattern” of market work? How do these characteristics of jobs and careers affect earnings?
12. This chapter considers the economics of discrimination by employers, customers, and governments. Now consider discrimination by workers. Suppose that some brunette workers did not like working with blonde workers. Do you think this worker discrimination could explain lower wages for blonde workers? If such a wage differential existed, what would a profit-maximizing entrepreneur do? If there were many such entrepreneurs, what would happen over time?

20



INCOME INEQUALITY AND POVERTY

“The only difference between the rich and other people,” Mary Colum once said to Ernest Hemingway, “is that the rich have more money.” Maybe so. But this claim leaves many questions unanswered. The gap between rich and poor is a fascinating and important topic of study—for the comfortable rich, for the struggling poor, and for the aspiring and worried middle class.

From the previous two chapters you should have some understanding about why different people have different incomes. A person’s earnings depend on the supply and demand for that person’s labor, which in turn depend on natural ability, human capital, compensating differentials, discrimination, and so on. Because labor earnings make up about three-fourths of the total income in the U.S. economy, the factors that determine wages are also largely responsible for determining how the economy’s total income is distributed among the various members of society. In other words, they determine who is rich and who is poor.

IN THIS CHAPTER
YOU WILL . . .

*Examine the degree
of economic
inequality in our
society*

*Consider some
problems that arise
when measuring
economic inequality*

*See how political
philosophers view
the government’s
role in
redistributing
income*

*Consider the
various policies
aimed at helping
poor families
escape poverty*



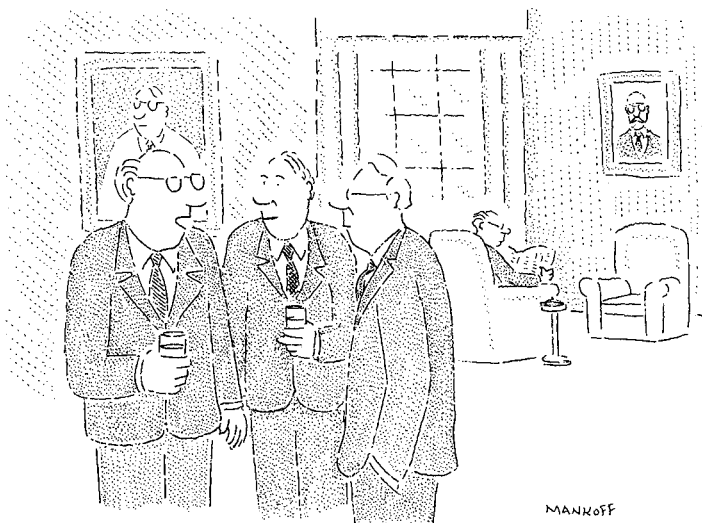
In this chapter we discuss the distribution of income. As we shall see, this topic raises some fundamental questions about the role of economic policy. One of the *Ten Principles of Economics* in Chapter 1 is that governments can sometimes improve market outcomes. This possibility is particularly important when considering the distribution of income. The invisible hand of the marketplace acts to allocate resources efficiently, but it does not necessarily ensure that resources are allocated fairly. As a result, many economists—though not all—believe that the government should redistribute income to achieve greater equality. In doing so, however, the government runs into another of the *Ten Principles of Economics*: People face tradeoffs. When the government enacts policies to make the distribution of income more equitable, it distorts incentives, alters behavior, and makes the allocation of resources less efficient.

Our discussion of the distribution of income proceeds in three steps. First, we assess how much inequality there is in our society. Second, we consider some different views about what role the government should play in altering the distribution of income. Third, we discuss various public policies aimed at helping society's poorest members.

THE MEASUREMENT OF INEQUALITY

We begin our study of the distribution of income by addressing four questions of measurement:

- ◆ How much inequality is there in our society?
- ◆ How many people live in poverty?



"As far as I'm concerned, they can do what they want with the minimum wage, just as long as they keep their hands off the maximum wage."

- ◆ What problems arise in measuring the amount of inequality?
- ◆ How often do people move among income classes?

These measurement questions are the natural starting point from which to discuss public policies aimed at changing the distribution of income.

U.S. INCOME INEQUALITY

There are various ways to describe the distribution of income in the economy. Table 20-1 presents a particularly simple way. It shows the percentage of families that fall into each of seven income categories. You can use this table to find where your family lies in the income distribution.

For examining differences in the income distribution over time or across countries, economists find it more useful to present the income data as in Table 20-2. To see how to interpret this table, consider the following thought experiment. Imagine that you lined up all the families in the economy according to their annual income. Then you divided the families into five equal groups: the bottom fifth, the second fifth, the middle fifth, the fourth fifth, and the top fifth. Next you computed the share of total income that each group of families received. In this way, you could produce the numbers in Table 20-2.

These numbers give us a way of gauging how the economy's total income is distributed. If income were equally distributed across all families, each one-fifth of families would receive one-fifth (20 percent) of income. If all income were concentrated among just a few families, the top fifth would receive 100 percent, and the other fifths would receive 0 percent. The actual economy, of course, is between these two extremes. The table shows that in 1998 the bottom fifth of all families received 4.2 percent of all income, and the top fifth of all families received 47.3 percent of all income. In other words, even though the top and bottom fifths include the same number of families, the top fifth has about ten times as much income as the bottom fifth.

The last column in Table 20-2 shows the share of total income received by the very richest families. In 1998, the top 5 percent of families received 20.7 percent of total income. Thus, the total income of the richest 5 percent of families was greater than the total income of the poorest 40 percent.

Table 20-2 also shows the distribution of income in various years beginning in 1935. At first glance, the distribution of income appears to have been remarkably stable over time. Throughout the past several decades, the bottom fifth of families

ANNUAL FAMILY INCOME	PERCENT OF FAMILIES
Less than \$15,000	11.7%
\$15,000-\$24,999	12.3
\$25,000-\$34,999	12.7
\$35,000-\$49,999	16.8
\$50,000-\$74,999	21.5
\$75,000-\$99,999	11.7
\$100,000 and over	13.3

Source: U.S. Bureau of the Census.

Table 20-1

THE DISTRIBUTION OF INCOME IN
THE UNITED STATES: 1998

Table 20-2

INCOME INEQUALITY IN THE UNITED STATES. This table shows the percent of total before-tax income received by families in each fifth of the income distribution and by those families in the top 5 percent.

YEAR	BOTTOM FIFTH	SECOND FIFTH	MIDDLE FIFTH	FOURTH FIFTH	TOP FIFTH	TOP 5 PERCENT
1998	4.2%	9.9%	15.7%	23.0%	47.3%	20.7%
1990	4.6	10.8	16.6	23.8	44.3	17.4
1980	5.2	11.5	17.5	24.3	41.5	15.3
1970	5.5	12.2	17.6	23.8	40.9	15.6
1960	4.8	12.2	17.8	24.0	41.3	15.9
1950	4.5	12.0	17.4	23.4	42.7	17.3
1935	4.1	9.2	14.1	20.9	51.7	26.5

Source: U.S. Bureau of the Census.

has received about 4 to 5 percent of income, while the top fifth has received about 40 to 50 percent of income. Closer inspection of the table reveals some trends in the degree of inequality. From 1935 to 1970, the distribution gradually became more equal. The share of the bottom fifth rose from 4.1 to 5.5 percent, and the share of the top fifth fell from 51.7 percent to 40.9 percent. In more recent years, this trend has reversed itself. From 1970 to 1998, the share of the bottom fifth fell from 5.5 percent to 4.2 percent, and the share of the top fifth rose from 40.9 to 47.3 percent.

In Chapter 19 we discussed some of the reasons for this recent increase in inequality. Increases in international trade with low-wage countries and changes in technology have tended to reduce the demand for unskilled labor and raise the demand for skilled labor. As a result, the wages of unskilled workers have fallen relative to the wages of skilled workers, and this change in relative wages has increased inequality in family incomes.

CASE STUDY THE WOMEN'S MOVEMENT AND THE INCOME DISTRIBUTION

Over the past several decades, there has been a dramatic change in women's role in the economy. The percentage of women who hold jobs has risen from about 32 percent in the 1950s to about 54 percent in the 1990s. As full-time homemakers have become less common, a woman's earnings have become a more important determinant of the total income of a typical family.

Although the women's movement has led to more equality between men and women in access to education and jobs, it has also led to less equality in family incomes. The reason is that the rise in women's labor-force participation has not been the same across all income groups. In particular, the women's movement has had its greatest impact on women from high-income households. Women from low-income households have long had high rates of participation in the labor force, even in the 1950s, and their behavior has changed much less.

In essence, the women's movement has changed the behavior of the wives of high-income men. In the 1950s, a male executive or physician was likely to marry a woman who would stay at home and raise the children. Today, the wife of a male executive or physician is more likely to be an executive or physician herself. The result is that rich households have become even richer, a pattern that raises inequality in family incomes.

As this example shows, there are social as well as economic determinants of the distribution of income. Moreover, the simplistic view that “income inequality is bad” can be misleading. Increasing the opportunities available to women was surely a good change for society, even if one effect was greater inequality in family incomes. When evaluating any change in the distribution of income, policymakers must look at the reasons for that change before deciding whether it presents a problem for society.



EQUALITY FOR WOMEN HAS MEANT
LESS EQUALITY FOR FAMILY INCOMES.

CASE STUDY INCOME INEQUALITY AROUND THE WORLD

How does the amount of income inequality in the United States compare to that in other countries? This question is interesting, but answering it is problematic. For many countries, data are not available. Even when they are, not every country in the world collects data in the same way; for example, some countries collect data on individual incomes, whereas other countries collect data on family incomes. As a result, whenever we find a difference between two countries, we can never be sure whether it reflects a true difference in the economies or merely a difference in the way data are collected.

With this warning in mind, consider Table 20-3, which compares the income distribution of the United States to that of seven other countries. The countries

COUNTRY	BOTTOM FIFTH	SECOND FIFTH	MIDDLE FIFTH	FOURTH FIFTH	TOP FIFTH
Germany	9.0%	13.5%	17.5%	22.9%	37.1%
Canada	7.5	12.9	17.2	23.0	39.3
Russia	7.4	12.6	17.7	24.2	38.2
United Kingdom	7.1	12.8	17.2	23.1	39.8
China	5.5	9.8	14.9	22.3	47.5
United States	4.8	10.5	16.0	23.5	45.2
Chile	3.5	6.6	10.9	18.1	61.0
Brazil	2.5	5.7	9.9	17.7	64.2

INCOME INEQUALITY AROUND THE WORLD. This table shows the percent of total before-tax income received by families in each fifth of the income distribution.

Table 20-3

Source: *World Development Report: 1998/99*, pp. 198–199.

are ranked from the most equal to the most unequal. On the top of the list is Germany, where the richest fifth of the population has income only about 4 times that of the poorest fifth. On the bottom of the list is Brazil, where the richest fifth has income about 25 times that of the poorest fifth. Although all countries have substantial inequality in income, the degree of inequality is not the same everywhere.

When countries are ranked by inequality, the United States ends up slightly behind the middle of the pack. Compared to other economically advanced countries, such as Germany and Canada, the United States has substantial inequality. But the United States has a more equal income distribution than many developing countries, such as Chile and Brazil.

THE POVERTY RATE

poverty rate

the percentage of the population whose family income falls below an absolute level called the poverty line

poverty line

an absolute level of income set by the federal government for each family size below which a family is deemed to be in poverty

A commonly used gauge of the distribution of income is the poverty rate. The **poverty rate** is the percentage of the population whose family income falls below an absolute level called the **poverty line**. The poverty line is set by the federal government at roughly three times the cost of providing an adequate diet. This line is adjusted every year to account for changes in the level of prices, and it depends on family size.

To get some idea about what the poverty rate tells us, consider the data for 1998. In that year, the median family had an income of \$47,469, and the poverty line for a family of four was \$16,660. The poverty rate was 12.7 percent. In other words, 12.7 percent of the population were members of families with incomes below the poverty line for their family size.

Figure 20-1 shows the poverty rate since 1959, when the official data begin. You can see that the poverty rate fell from 22.4 percent in 1959 to a low of 11.1 percent in 1973. This decline is not surprising, for average income in the economy (adjusted

Figure 20-1

THE POVERTY RATE. The poverty rate shows the percentage of the population with incomes below an absolute level called the poverty line.

Source: U.S. Bureau of the Census.

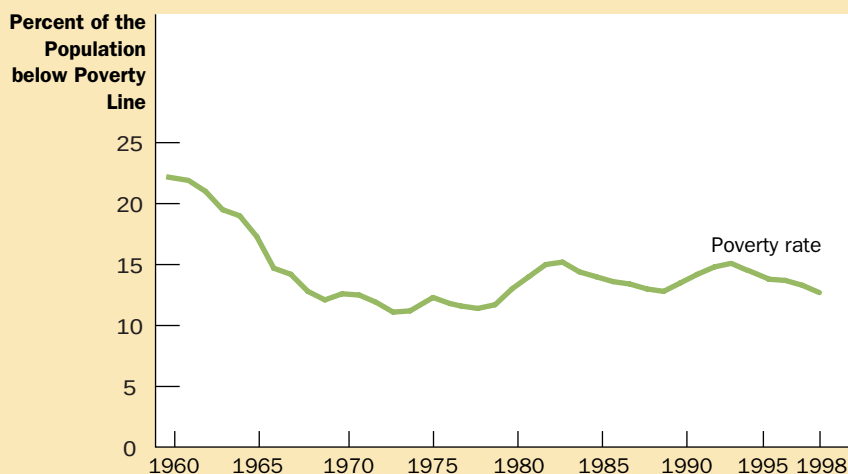


Table 20-4

GROUP	POVERTY RATE
All persons	12.7%
White, not Hispanic	8.2
Black	26.1
Hispanic	25.6
Asian, Pacific Islander	12.5
Children (under age 18)	18.9
Elderly (over age 64)	10.5
Female household, no husband present	33.1

Source: U.S. Bureau of the Census. Data are for 1998.

WHO IS POOR? This table shows that the poverty rate varies greatly among different groups within the population.

for inflation) rose more than 50 percent during this period. Because the poverty line is an absolute rather than a relative standard, more families are pushed above the poverty line as economic growth pushes the entire income distribution upward. As John F. Kennedy once put it, a rising tide lifts all boats.

Since the early 1970s, however, the economy's rising tide has left some boats behind. Despite continued (although somewhat slower) growth in average income, the poverty rate has not declined. This lack of progress in reducing poverty in recent years is closely related to the increasing inequality we saw in Table 20-2. Although economic growth has raised the income of the typical family, the increase in inequality has prevented the poorest families from sharing in this greater economic prosperity.

Poverty is an economic malady that affects all groups within the population, but it does not affect all groups with equal frequency. Table 20-4 shows the poverty rates for several groups, and it reveals three striking facts:

- ◆ Poverty is correlated with race. Blacks and Hispanics are about three times more likely to live in poverty than are whites.
- ◆ Poverty is correlated with age. Children are more likely than average to be members of poor families, and the elderly are less likely than average to be poor.
- ◆ Poverty is correlated with family composition. Families headed by a female adult and without a husband present are more than twice as likely to live in poverty as the average family.

These three facts have described U.S. society for many years, and they show which people are most likely to be poor. These effects also work together: Among black and Hispanic children in female-headed households, more than half live in poverty.

PROBLEMS IN MEASURING INEQUALITY

Although data on the income distribution and the poverty rate help to give us some idea about the degree of inequality in our society, interpreting these data is not as straightforward as it might first appear. The data are based on households'

annual incomes. What people care about, however, is not their incomes but their ability to maintain a good standard of living. For various reasons, data on the income distribution and the poverty rate give an incomplete picture of inequality in living standards. We examine these reasons below.

in-kind transfers

transfers to the poor given in the form of goods and services rather than cash

In-Kind Transfers Measurements of the distribution of income and the poverty rate are based on families' money income. Through various government programs, however, the poor receive many nonmonetary items, including food stamps, housing vouchers, and medical services. Transfers to the poor given in the form of goods and services rather than cash are called **in-kind transfers**. Standard measurements of the degree of inequality do not take account of these in-kind transfers.

Because in-kind transfers are received mostly by the poorest members of society, the failure to include in-kind transfers as part of income greatly affects the measured poverty rate. According to a study by the Census Bureau, if in-kind transfers were included in income at their market value, the number of families in poverty would be about 10 percent lower than the standard data indicate.

The important role of in-kind transfers makes evaluating changes in poverty more difficult. Over time, as public policies to help the poor evolve, the composition of assistance between cash and in-kind transfers changes. Some of the fluctuations in the measured poverty rate, therefore, reflect the form of government assistance rather than the true extent of economic deprivation.

life cycle

the regular pattern of income variation over a person's life

The Economic Life Cycle Incomes vary predictably over people's lives. A young worker, especially one in school, has a low income. Income rises as the worker gains maturity and experience, peaks at around age 50, and then falls sharply when the worker retires at around age 65. This regular pattern of income variation is called the **life cycle**.

Because people can borrow and save to smooth out life cycle changes in income, their standard of living in any year depends more on lifetime income than on that year's income. The young often borrow, perhaps to go to school or to buy a house, and then repay these loans later when their incomes rise. People have their highest saving rates when they are middle-aged. Because people can save in anticipation of retirement, the large declines in incomes at retirement need not lead to similar declines in the standard of living.

This normal life cycle pattern causes inequality in the distribution of annual income, but it does not represent true inequality in living standards. To gauge the inequality of living standards in our society, the distribution of lifetime incomes is more relevant than the distribution of annual incomes. Unfortunately, data on lifetime incomes are not readily available. When looking at any data on inequality, however, it is important to keep the life cycle in mind. Because a person's lifetime income smooths out the highs and lows of the life cycle, lifetime incomes are surely more equally distributed across the population than are annual incomes.

Transitory versus Permanent Income Incomes vary over people's lives not only because of predictable life cycle variation but also because of random and transitory forces. One year a frost kills off the Florida orange crop, and Florida orange growers see their incomes fall temporarily. At the same time, the

IN THE NEWS

Measuring Poverty

HOW MANY PEOPLE LIVE IN POVERTY? THE answer is a topic of continuing debate.

Devising New Math to Define Poverty

BY LOUIS UCHITELLE

The Census Bureau has begun to revise its definition of what constitutes poverty in the United States, experimenting with a formula that would drop millions of more families below the poverty line.

The bureau's new approach would in effect raise the income threshold for living above poverty to \$19,500 for a

family of four, from the \$16,600 now considered sufficient. Suddenly, 46 million Americans, or 17 percent of the population, would be recognized as officially below the line, not the 12.7 percent announced last month, the lowest in nearly a decade. . . .

Fixing a poverty line has always been a subjective endeavor. The current formula was created for President Lyndon B. Johnson to keep score on his "war on poverty" and has remained unchanged since 1965 except for adjustments for inflation. . . . The Census Bureau's new Experimental Measures are an effort to determine what poor people must spend on food, clothing, housing, and life's little extras.

"There is no scientific way to set a new poverty line," said Rebecca M. Blank, dean of the School of Social Policy at the University of Michigan. "What there is here are a set of judgment calls, now being made, about what is needed to lift people to a socially acceptable standard of living." . . .

Ordinary Americans, in opinion polls, draw the poverty line above \$20,000, saying it takes at least that much, if not more, to "get along in their community," to "live decently," or to avoid hardship.

But a higher threshold means government spending would rise to pay for benefits tied to the poverty level, like food stamps and Head Start. That would require an incursion into the budget surplus that neither Republicans nor Democrats seek.

Not surprising, the White House, which would have to authorize a change in the poverty formula, is proceeding cautiously. "We have at least a couple of years more work to do," an Administration official said, passing the decision for redefining poverty to the next administration.

SOURCE: *The New York Times*, October 18, 1999, pp. A1, A14.

Florida frost drives up the price of oranges, and California orange growers see their incomes temporarily rise. The next year the reverse might happen.

Just as people can borrow and lend to smooth out life cycle variation in income, they can also borrow and lend to smooth out transitory variation in income. When California orange growers experience a good year, they would be foolish to spend all of their additional income. Instead, they save some of it, knowing that their good fortune is unlikely to persist. Similarly, the Florida orange growers respond to their temporarily low incomes by drawing down their savings or by borrowing. To the extent that a family saves and borrows to buffer itself from transitory changes in income, these changes do not affect its standard of living. A family's ability to buy goods and services depends largely on its **permanent income**, which is its normal, or average, income.

To gauge inequality of living standards, the distribution of permanent income is more relevant than the distribution of annual income. Although permanent income is hard to measure, it is an important concept. Because permanent income excludes transitory changes in income, permanent income is more equally distributed than is current income.

permanent income
a person's normal income

ECONOMIC MOBILITY

People sometimes speak of “the rich” and “the poor” as if these groups consisted of the same families year after year. In fact, this is not at all the case. Economic mobility, the movement of people among income classes, is substantial in the U.S. economy. Movements up the income ladder can be due to good luck or hard work, and movements down the ladder can be due to bad luck or laziness. Some of this mobility reflects transitory variation in income, while some reflects more persistent changes in income.

Because economic mobility is so great, many of those below the poverty line are there only temporarily. Poverty is a long-term problem for relatively few families. In a typical ten-year period, about one in four families falls below the poverty line in at least one year. Yet fewer than 3 percent of families are poor for eight or more years. Because it is likely that the temporarily poor and the persistently poor face different problems, policies that aim to combat poverty need to distinguish between these groups.

Another way to gauge economic mobility is the persistence of economic success from generation to generation. Economists who have studied this topic find substantial mobility. If a father earns 20 percent above his generation’s average income, his son will most likely earn 8 percent above his generation’s average income. There is almost no correlation between the income of a grandfather and the income of a grandson. There is much truth to the old saying, “From shirtsleeves to shirtsleeves in three generations.”

One result of this great economic mobility is that the U.S. economy is filled with self-made millionaires (as well as with heirs who squandered the fortunes they inherited). According to estimates for 1996, about 2.7 million households in the United States had net worth (assets minus debts) that exceeded \$1 million. These households represented the richest 2.8 percent of the population. About four out of five of these millionaires made their money on their own, such as by starting and building a business or by climbing the corporate ladder. Only one in five millionaires inherited their fortunes.

QUICK QUIZ: What does the poverty rate measure? ♦ Describe three potential problems in interpreting the measured poverty rate.

THE POLITICAL PHILOSOPHY OF REDISTRIBUTING INCOME

We have just seen how the economy’s income is distributed and have considered some of the problems in interpreting measured inequality. This discussion was *positive* in the sense that it merely described the world as it is. We now turn to the *normative* question facing policymakers: What should the government do about economic inequality?

This question is not just about economics. Economic analysis alone cannot tell us whether policymakers should try to make our society more egalitarian. Our views on this question are, to a large extent, a matter of political philosophy. Yet

because the government's role in redistributing income is central to so many debates over economic policy, here we digress from economic science to consider a bit of political philosophy.

UTILITARIANISM

A prominent school of thought in political philosophy is **utilitarianism**. The founders of utilitarianism are the English philosophers Jeremy Bentham (1748–1832) and John Stuart Mill (1806–1873). To a large extent, the goal of utilitarians is to apply the logic of individual decisionmaking to questions concerning morality and public policy.

The starting point of utilitarianism is the notion of **utility**—the level of happiness or satisfaction that a person receives from his or her circumstances. Utility is a measure of well-being and, according to utilitarians, is the ultimate objective of all public and private actions. The proper goal of the government, they claim, is to maximize the sum of utility of everyone in society.

The utilitarian case for redistributing income is based on the assumption of *diminishing marginal utility*. It seems reasonable that an extra dollar of income to a poor person provides that person with more additional utility than does an extra dollar to a rich person. In other words, as a person's income rises, the extra well-being derived from an additional dollar of income falls. This plausible assumption, together with the utilitarian goal of maximizing total utility, implies that the government should try to achieve a more equal distribution of income.

The argument is simple. Imagine that Peter and Paul are the same, except that Peter earns \$80,000 and Paul earns \$20,000. In this case, taking a dollar from Peter to pay Paul will reduce Peter's utility and raise Paul's utility. But, because of diminishing marginal utility, Peter's utility falls by less than Paul's utility rises. Thus, this redistribution of income raises total utility, which is the utilitarian's objective.

At first, this utilitarian argument might seem to imply that the government should continue to redistribute income until everyone in society has exactly the same income. Indeed, that would be the case if the total amount of income—\$100,000 in our example—were fixed. But, in fact, it is not. Utilitarians reject complete equalization of incomes because they accept one of the *Ten Principles of Economics* presented in Chapter 1: People respond to incentives.

To take from Peter to pay Paul, the government must pursue policies that redistribute income, such as the U.S. federal income tax and welfare system. Under these policies, people with high incomes pay high taxes, and people with low incomes receive income transfers. Yet, as we have seen in Chapters 8 and 12, taxes distort incentives and cause deadweight losses. If the government takes away additional income a person might earn through higher income taxes or reduced transfers, both Peter and Paul have less incentive to work hard. As they work less, society's income falls, and so does total utility. The utilitarian government has to balance the gains from greater equality against the losses from distorted incentives. To maximize total utility, therefore, the government stops short of making society fully egalitarian.

A famous parable sheds light on the utilitarian's logic. Imagine that Peter and Paul are thirsty travelers trapped at different places in the desert. Peter's oasis has much water; Paul's has little. If the government could transfer water from one oasis

utilitarianism

the political philosophy according to which the government should choose policies to maximize the total utility of everyone in society

utility

a measure of happiness or satisfaction



to the other without cost, it would maximize total utility from water by equalizing the amount in the two places. But suppose that the government has only a leaky bucket. As it tries to move water from one place to the other, some of the water is lost in transit. In this case, a utilitarian government might still try to move some water from Peter to Paul, depending on how thirsty Paul is and how leaky the bucket is. But, with only a leaky bucket at its disposal, a utilitarian government will not try to reach complete equality.

LIBERALISM

liberalism

the political philosophy according to which the government should choose policies deemed to be just, as evaluated by an impartial observer behind a “veil of ignorance”

A second way of thinking about inequality might be called **liberalism**. Philosopher John Rawls develops this view in his book *A Theory of Justice*. This book was first published in 1971, and it quickly became a classic in political philosophy.

Rawls begins with the premise that a society’s institutions, laws, and policies should be just. He then takes up the natural question: How can we, the members of society, ever agree on what justice means? It might seem that every person’s point of view is inevitably based on his or her particular circumstances—whether he or she is talented or less talented, diligent or lazy, educated or less educated, born to a wealthy family or a poor one. Could we ever *objectively* determine what a just society would be?

To answer this question, Rawls proposes the following thought experiment. Imagine that before any of us is born, we all get together for a meeting to design the rules that govern society. At this point, we are all ignorant about the station in life each of us will end up filling. In Rawls’s words, we are sitting in an “original position” behind a “veil of ignorance.” In this original position, Rawls argues, we can choose a just set of rules for society because we must consider how those rules will affect every person. As Rawls puts it, “Since all are similarly situated and no one is able to design principles to favor his particular conditions, the principles of justice are the result of fair agreement or bargain.” Designing public policies and institutions in this way allows us to be objective about what policies are just.

Rawls then considers what public policy designed behind this veil of ignorance would try to achieve. In particular, he considers what income distribution a person would consider just if that person did not know whether he or she would end up at the top, bottom, or middle of the distribution. Rawls argues that a person in the original position would be especially concerned about the possibility of being at the *bottom* of the income distribution. In designing public policies, therefore, we should aim to raise the welfare of the worst-off person in society. That is, rather than maximizing the sum of everyone’s utility, as a utilitarian would do, Rawls would maximize the minimum utility. Rawls’s rule is called the **maximin criterion**.

maximin criterion

the claim that the government should aim to maximize the well-being of the worst-off person in society

Because the maximin criterion emphasizes the least fortunate person in society, it justifies public policies aimed at equalizing the distribution of income. By transferring income from the rich to the poor, society raises the well-being of the least fortunate. The maximin criterion would not, however, lead to a completely egalitarian society. If the government promised to equalize incomes completely, people would have no incentive to work hard, society’s total income would fall substantially, and the least fortunate person would be worse off. Thus, the maximin criterion still allows disparities in income, because such disparities can improve incentives and thereby raise society’s ability to help the poor. Nonetheless,

IN THE NEWS

A Rawlsian Billionaire

INVESTOR WARREN BUFFETT'S \$36 BILLION make him one of the world's richest men. Here is how Buffett explained his personal philosophy to an audience of college students at the University of Washington. He is responding to a question about the importance of "giving back to your community." Notice the echoes of Rawls's veil of ignorance.

Buffett's Answer

I know in my own case that 99%-plus [of my wealth] will go back to society, just because we've been treated extraordinarily well by society.

I'm lucky. I don't run fast, but I'm wired in a particular way that I thrive in a big capitalist economy with a lot of action. . . . If I had been born some time ago I would've been some animal's lunch. . . .

Let me suggest another way to think about this. Let's say that it was 24 hours before you were born, and a genie appeared and said, "You look like a winner. I have enormous confidence in you, and what I'm going to do is let you set the rules for society into which you will be born. You can set the economic rules and the social rules, and whatever rules

you set will apply during your lifetime and your children's lifetime."

And you'll say, "Well, that's nice, but what's the catch?"

And the genie says, "Here's the catch. You don't know if you're going to be born rich or poor, black or white, male or female, able-bodied or infirm, intelligent or retarded." So all you know is that you're going to get one ball out of a barrel with, say, 5.8 billion balls in it [each ball representing one of the 5.8 billion people on earth]. You're going to participate in what I call the ovarian lottery. It's the most important thing that will happen to you in your life, but you have no control over it. It's going to determine far more than your grades at school or anything else that happens to you.

Now, what rules do you want to have? I'm not going to tell you the rules, and nobody will tell you; you have to make them up for yourself. But they will affect how you think about what you do in your will and things of that sort.

You're going to want to have a system that turns out more and more goods and services. You've got a great quantity of people out there, and you want them to live pretty well, and you want your kids to live better than you did, and you want your grandchildren to live better than your kids. You're going to want a system that keeps Bill Gates and Andy Grove and Jack Welch [heads of Microsoft, Intel, and General Electric] working long, long after they don't need to work. You're going to want the most able people working more than 12 hours a day. So you've got to have a system that gives them an incentive to turn out the goods and services.

But you're also going to want a system that takes care of the bad balls, the ones that aren't lucky. If you have a



WARREN BUFFETT

system that is turning out enough goods and services, you can take care of them. You don't want people worrying about being sick in their old age, or fearful about going home at night. You want a system where people are free of fear to some extent.

So you'll try to design something, assuming you have the goods and services to solve that sort of thing. You'll want equality of opportunity—namely a good public school system—to make you feel that every piece of talent out there will get the same shot at contributing. And your tax system will follow from your reasoning on that. And what you do with the money you make is another thing to think about. As you work through that, everybody comes up with something a little different. I just suggest that you play that little game.

SOURCE: *Fortune*, July 20, 1998, pp. 62–64.

because Rawls's philosophy puts weight on only the least fortunate members of society, it calls for more income redistribution than does utilitarianism.

Rawls's views are controversial, but the thought experiment he proposes has much appeal. In particular, this thought experiment allows us to consider the redistribution of income as a form of *social insurance*. That is, from the perspective of the original position behind the veil of ignorance, income redistribution is like an insurance policy. Homeowners buy fire insurance to protect themselves from the risk of their housing burning down. Similarly, when we as a society choose policies that tax the rich to supplement the incomes of the poor, we are all insuring ourselves against the possibility that we might have been a member of a poor family. Because people dislike risk, we should be happy to have been born into a society that provides us this insurance.

It is not at all clear, however, that rational people behind the veil of ignorance would truly be so averse to risk as to follow the maximin criterion. Indeed, because a person in the original position might end up anywhere in the distribution of outcomes, he or she might treat all possible outcomes equally when designing public policies. In this case, the best policy behind the veil of ignorance would be to maximize the average utility of members of society, and the resulting notion of justice would be more utilitarian than Rawlsian.

LIBERTARIANISM

libertarianism

the political philosophy according to which the government should punish crimes and enforce voluntary agreements but not redistribute income

A third view of inequality is called **libertarianism**. The two views we have considered so far—utilitarianism and liberalism—both view the total income of society as a shared resource that a social planner can freely redistribute to achieve some social goal. By contrast, libertarians argue that society itself earns no income—only individual members of society earn income. According to libertarians, the government should not take from some individuals and give to others in order to achieve any particular distribution of income.

For instance, philosopher Robert Nozick writes the following in his famous 1974 book *Anarchy, State and Utopia*:

We are not in the position of children who have been given portions of pie by someone who now makes last minute adjustments to rectify careless cutting. There is no *central* distribution, no person or group entitled to control all the resources, jointly deciding how they are to be doled out. What each person gets, he gets from others who give to him in exchange for something, or as a gift. In a free society, diverse persons control different resources, and new holdings arise out of the voluntary exchanges and actions of persons.

Whereas utilitarians and liberals try to judge what amount of inequality is desirable in a society, Nozick denies the validity of this very question.

The libertarian alternative to evaluating economic *outcomes* is to evaluate the *process* by which these outcomes arise. When the distribution of income is achieved unfairly—for instance, when one person steals from another—the government has the right and duty to remedy the problem. But, as long as the process determining the distribution of income is just, the resulting distribution is fair, no matter how unequal.

Nozick criticizes Rawls's liberalism by drawing an analogy between the distribution of income in society and the distribution of grades in a course. Suppose you were asked to judge the fairness of the grades in the economics course you are

now taking. Would you imagine yourself behind a veil of ignorance and choose a grade distribution without knowing the talents and efforts of each student? Or would you ensure that the process of assigning grades to students is fair without regard for whether the resulting distribution is equal or unequal? For the case of grades at least, the libertarian emphasis on process over outcomes is compelling.

Libertarians conclude that equality of opportunities is more important than equality of incomes. They believe that the government should enforce individual rights to ensure that everyone has the same opportunity to use his or her talents and achieve success. Once these rules of the game are established, the government has no reason to alter the resulting distribution of income.

QUICK QUIZ: Pam earns more than Pauline. Someone proposes taxing Pam in order to supplement Pauline's income. How would a utilitarian, a liberal, and a libertarian evaluate this proposal?

POLICIES TO REDUCE POVERTY

As we have just seen, political philosophers hold various views about what role the government should take in altering the distribution of income. Political debate among the larger population of voters reflects a similar disagreement. Despite these continuing debates, however, most people believe that, at the very least, the government should try to help those most in need. According to a popular metaphor, the government should provide a "safety net" to prevent any citizen from falling too far.

Poverty is one of the most difficult problems that policymakers face. Poor families are more likely than the overall population to experience homelessness, drug dependency, domestic violence, health problems, teenage pregnancy, illiteracy, unemployment, and low educational attainment. Members of poor families are both more likely to commit crimes and more likely to be victims of crimes. Although it is hard to separate the causes of poverty from the effects, there is no doubt that poverty is associated with various economic and social ills.

Suppose that you were a policymaker in the government, and your goal was to reduce the number of people living in poverty. How would you achieve this goal? Here we consider some of the policy options that you might consider. Although each of these options does help some people escape poverty, none of them is perfect, and deciding which is best is not easy.

MINIMUM-WAGE LAWS

Laws setting a minimum wage that employers can pay workers are a perennial source of debate. Advocates view the minimum wage as a way of helping the working poor without any cost to the government. Critics view it as hurting those it is intended to help.

The minimum wage is easily understood using the tools of supply and demand, as we first saw in Chapter 6. For workers with low levels of skill and experience, a

high minimum wage forces the wage above the level that balances supply and demand. It therefore raises the cost of labor to firms and reduces the quantity of labor that those firms demand. The result is higher unemployment among those groups of workers affected by the minimum wage. Although those workers who remain employed benefit from a higher wage, those who might have been employed at a lower wage are worse off.

The magnitude of these effects depends crucially on the elasticity of demand. Advocates of a high minimum wage argue that the demand for unskilled labor is relatively inelastic, so that a high minimum wage depresses employment only slightly. Critics of the minimum wage argue that labor demand is more elastic, especially in the long run when firms can adjust employment and production more fully. They also note that many minimum-wage workers are teenagers from middle-class families, so that a high minimum wage is imperfectly targeted as a policy for helping the poor.

WELFARE

welfare

government programs that supplement the incomes of the needy

One way to raise the living standards of the poor is for the government to supplement their incomes. The primary way in which the government does this is through the welfare system. **Welfare** is a broad term that encompasses various government programs. Temporary Assistance for Needy Families (formerly called Aid to Families with Dependent Children) is a program that assists families where there are children but no adult able to support the family. In a typical family receiving such assistance, the father is absent, and the mother is at home raising small children. Another welfare program is Supplemental Security Income (SSI), which provides assistance to the poor who are sick or disabled. Note that for both of these welfare programs, a poor person cannot qualify for assistance simply by having a low income. He or she must also establish some additional “need,” such as small children or a disability.

A common criticism of welfare programs is that they create incentives for people to become “needy.” For example, these programs may encourage families to break up, for many families qualify for financial assistance only if the father is absent. The programs may also encourage illegitimate births, for many poor, single women qualify for assistance only if they have children. Because poor, single mothers are such an important part of the poverty problem and because welfare programs seem to raise the number of poor, single mothers, critics of the welfare system assert that these policies exacerbate the very problems they are supposed to cure. As a result of these arguments, the welfare system was revised in a 1996 law that limited the amount of time recipients could stay on welfare.

How severe are these potential problems with the welfare system? No one knows for sure. Proponents of the welfare system point out that being a poor, single mother on welfare is a difficult existence at best, and they are skeptical that many people would be encouraged to pursue such a life if it were not thrust upon them. Moreover, trends over time do not support the view that the decline of the two-parent family is largely a symptom of the welfare system, as the system’s critics sometimes claim. Since the early 1970s, welfare benefits (adjusted for inflation) have declined, yet the percentage of children living with only one parent has risen.

IN THE NEWS*Should the Government Try to Help Poor Regions?*

MANY ANTIPOVERTY PROGRAMS ARE TARGETED at poor areas of the country. Economist Edward Glaeser presents the case against this geographic approach.

**Help Poor People,
Not Poor Places**

BY EDWARD L. GLAESER

President Clinton's six-city "New Markets" tour earlier this summer signaled a renewed focus on the problems of the poor. But while the president's concern is appreciated by all of us who care about the islands of poverty in America's sea of affluence, his proposals are fundamentally flawed. They may still help some of the poor, but also risk repeating some of the worst mistakes of the Johnson era.

The trouble with the president's recommendations is that they violate the first economic rule of urban poverty policy: Programs should be person-based, not place-based.

Economists have long argued that place-based programs are a mistake. They strongly prefer person-based policies that create transfers, entitlements, or relief from regulation on the basis of personal characteristics. Examples of person-based policies include the Earned Income Tax Credit and the GI Bill.

Place-based policies, on the other hand, give transfers or other government support on the basis of location. Examples of such policies are housing projects and enterprise zones. President Clinton's recent Rural Housing and Economic Development Assistance for Kentucky or the new Empowerment Zone Grant for East St. Louis, Ill., are quintessential place-based policies.

The problem with place-based programs is that they create incentives to keep the poor in the ghetto. By subsidizing the place, not the person living there, these policies make it more attractive for the poor to stay in high-poverty areas. Indeed, current research shows that supposedly benevolent pro-poor housing and transfer policies play a major role in herding the poor into inner cities.

It's hard to see the logic in artificially limiting migration and concentrating the poor in areas with low productivity. Movement out of low-productivity, high-unemployment areas is one reason that unemployment rates in the U.S. stay low. Moreover, flight from the ghettos

has enabled many African-Americans to avoid the social costs of the inner city, and black-white segregation in the U.S. has declined substantially because of it.

Place-based programs also suffer from the fact that their benefits go disproportionately to property owners in the targeting areas—and not to the intended beneficiaries. If the government offers tax credits to firms that invest in a poor region, for instance, then firms will locate there, pushing up property values and rents. But the benefits of increased economic activity will evaporate as higher housing costs eat away the planned benefits to the needy. . . .

If place-based policies are so bad, why are they so popular? A cynic might say that the residents of wealthy suburbs prefer that the poor remain in ghettos. A more practical explanation is that we have place-based politicians who lobby for place-based policies. . . .

A wise alternative to such faulty place-based poverty assistance would be a program that offers tax credits to companies that employ the disadvantaged. This would be a less distortionary means of assisting the poor.

SOURCE: *The Wall Street Journal*, August 12, 1999, p. A22.

NEGATIVE INCOME TAX

Whenever the government chooses a system to collect taxes, it affects the distribution of income. This is clearly true in the case of a progressive income tax, whereby high-income families pay a larger percentage of their income in taxes than do low-income families. As we discussed in Chapter 12, equity across income groups is an important criterion in the design of a tax system.

negative income tax

a tax system that collects revenue from high-income households and gives transfers to low-income households

Many economists have advocated supplementing the income of the poor using a **negative income tax**. According to this policy, every family would report its income to the government. High-income families would pay a tax based on their incomes. Low-income families would receive a subsidy. In other words, they would “pay” a “negative tax.”

For example, suppose the government used the following formula to compute a family’s tax liability:

$$\text{Taxes owed} = (1/3 \text{ of income}) - \$10,000.$$

In this case, a family that earned \$60,000 would pay \$10,000 in taxes, and a family that earned \$90,000 would pay \$20,000 in taxes. A family that earned \$30,000 would owe nothing. And a family that earned \$15,000 would “owe” –\$5,000. In other words, the government would send this family a check for \$5,000.

Under a negative income tax, poor families would receive financial assistance without having to demonstrate need. The only qualification required to receive assistance would be a low income. Depending on one’s point of view, this feature can be either an advantage or a disadvantage. On the one hand, a negative income tax does not encourage illegitimate births and the breakup of families, as critics of the welfare system believe current policy does. On the other hand, a negative income tax would subsidize those who are simply lazy and, in some people’s eyes, undeserving of government support.

One actual tax provision that works much like a negative income tax is the Earned Income Tax Credit. This credit allows poor working families to receive income tax refunds greater than the taxes they paid during the year. Because the Earned Income Tax Credit applies only to the working poor, it does not discourage recipients from working, as other antipoverty programs are claimed to do. For the same reason, however, it also does not help alleviate poverty due to unemployment, sickness, or other inability to work.

IN-KIND TRANSFERS

Another way to help the poor is to provide them directly with some of the goods and services they need to raise their living standards. For example, charities provide the needy with food, shelter, and toys at Christmas. The government gives poor families *food stamps*, which are government vouchers that can be used to buy food at stores; the stores then redeem the vouchers for money. The government also gives many poor people health care through a program called *Medicaid*.

Is it better to help the poor with these in-kind transfers or with direct cash payments? There is no clear answer.

Advocates of in-kind transfers argue that such transfers ensure that the poor get what they need most. Among the poorest members of society, alcohol and drug addiction is more common than it is in society as a whole. By providing the poor with food and shelter, society can be more confident that it is not helping to support such addictions. This is one reason why in-kind transfers are more politically popular than cash payments to the poor.

Advocates of cash payments argue that in-kind transfers are inefficient and disrespectful. The government does not know what goods and services the poor need most. Many of the poor are ordinary people down on their luck. Despite their

misfortune, they are in the best position to decide how to raise their own living standards. Rather than giving the poor in-kind transfers of goods and services that they may not want, it may be better to give them cash and allow them to buy what they think they need most.

ANTIPOVERTY PROGRAMS AND WORK INCENTIVES

Many policies aimed at helping the poor can have the unintended effect of discouraging the poor from escaping poverty on their own. To see why, consider the following example. Suppose that a family needs an income of \$15,000 to maintain a reasonable standard of living. And suppose that, out of concern for the poor, the government promises to guarantee every family that income. Whatever a family earns, the government makes up the difference between that income and \$15,000. What effect would you expect this policy to have?

The incentive effects of this policy are obvious: Any person who would make under \$15,000 by working has no incentive to find and keep a job. For every dollar that the person would earn, the government would reduce the income supplement by a dollar. In effect, the government taxes 100 percent of additional earnings. An effective marginal tax rate of 100 percent is surely a policy with a large dead-weight loss.

The adverse effects of this high effective tax rate can persist over time. A person discouraged from working loses the on-the-job training that a job might offer. In addition, his or her children miss the lessons learned by observing a parent with a full-time job, and this may adversely affect their own ability to find and hold a job.

Although the antipoverty program we have been discussing is hypothetical, it is not as unrealistic as it might first appear. Welfare, Medicaid, food stamps, and the Earned Income Tax Credit are all programs aimed at helping the poor, and they are all tied to family income. As a family's income rises, the family becomes ineligible for these programs. When all these programs are taken together, it is common for families to face effective marginal tax rates that are very high. Sometimes the effective marginal tax rates even exceed 100 percent, so that poor families are worse off when they earn more. By trying to help the poor, the government discourages those families from working. According to critics of antipoverty programs, these programs alter work attitudes and create a "culture of poverty."

It might seem that there is an easy solution to this problem: Reduce benefits to poor families more gradually as their incomes rise. For example, if a poor family loses 30 cents of benefits for every dollar it earns, then it faces an effective marginal tax rate of 30 percent. Although this effective tax reduces work effort to some extent, it does not eliminate the incentive to work completely.

The problem with this solution is that it greatly increases the cost of programs to combat poverty. If benefits are phased out gradually as a poor family's income rises, then families just above the poverty level will also be eligible for substantial benefits. The more gradual the phase-out, the more families are eligible, and the greater the cost of the program. Thus, policymakers face a tradeoff between burdening the poor with high effective marginal tax rates and burdening taxpayers with costly programs to reduce poverty.

There are various other ways to try to reduce the work disincentive of antipoverty programs. One is to require any person collecting benefits to accept a

IN THE NEWS*Welfare Reform*

IN 1996 THE U.S. WELFARE SYSTEM UNDERWENT a major reform, including the enactment of time limits on benefits. In the following opinion column, economist Gary Becker evaluates the change in policy.

Guess What?
Welfare Reform Works

BY GARY S. BECKER

The welfare reform act of 1996 is one of the most revolutionary pieces of legislation since the welfare state began half a

century ago. Contrary to the predictions of many skeptics, this law has been remarkably successful—helped, to be sure, by the strong economy of the past several years.

The success of earlier reforms by a few states led to a bipartisan effort by a Republican Congress and President Clinton to “end welfare as we know it” by forcing recipients in all states to look for work. The 1996 law limits families to two years of welfare income during any one spell and caps the total time on welfare over a mother’s lifetime at five years.

The number of recipients rose sharply from the early 1960s, to a peak in 1993, when over 4 million American families were on welfare. In that year, an incredible 1 million residents of New York City alone were receiving welfare, up from 250,000 in 1960.

Wisconsin, Massachusetts, New Jersey, and a few other states decided in the late 1980s that this upward trend in

welfare was unacceptable and could be reversed. They introduced reforms that discouraged women from having children while on welfare. More important, they dropped the assumption that most women on welfare are not capable of getting and holding jobs, and put pressure on them to find employment to help support their families.

These states managed to cut their welfare populations while at the same time improving the economic situation of single-parent families. Recently, a careful evaluation of the Massachusetts reforms by economists M. Anne Hill and Thomas J. Main of the City University of New York concluded that they not only greatly reduced that state’s welfare caseload but also encouraged more young women to finish high school and sharpen their economic skills.

What worked in Massachusetts and other pioneering states was applicable throughout the nation, but the reassessment of federal welfare policy

government-provided job—a system sometimes called *workfare*. Another possibility is to provide benefits for only a limited period of time. This route was taken in a 1996 welfare reform bill. Advocates of time limits point to the falling poverty rate in the late 1990s as evidence supporting this approach. Critics argue that time limits are cruel to the least fortunate members of society and that the falling poverty rate in the late 1990s is due more to a strong economy than to welfare reform.

QUICK QUIZ: List three policies aimed at helping the poor, and discuss the pros and cons of each.

CONCLUSION

People have long reflected on the distribution of income in society. Plato, the ancient Greek philosopher, concluded that in an ideal society the income of the richest

was opposed by many intellectuals. Some members of President Clinton's team quit after the 1996 federal law, over what they considered a betrayal of the welfare state. They argued that most women forced off welfare would become homeless or destitute, since they supposedly are too mentally or physically handicapped or lacking in requisite skills to obtain and hold jobs.

However, this law has been highly successful in reversing the large growth in the number of welfare recipients in the United States. Most mothers forced off welfare found work and provide financial support for their children.

Certainly, the huge decline—by over 40 percent—in the number of single mothers on welfare from the 1993 peak is partly due to the booming economy of the past seven years. However, most of this decline took place in the two years after the passage of the 1996 act. The study of Massachusetts' experience cited earlier confirms the importance of

the new approach to welfare, since the authors' research attributes more than one-third of the decline in that state's welfare role since 1995 to the reforms and not simply to its buoyant economy.

The federal law recognizes that the number of families in need of assistance always rises sharply during bad economic times. This is why each welfare spell is allowed to last up to two years, and mothers with dependent children can have multiple spells, up to a total of five years over their lifetimes. It further acknowledges that some women are handicapped and unable to work. What it aims to discourage is the attraction of welfare to able-bodied women during good times when jobs are available.

The act also recognizes that many poor working mothers will not earn enough to provide a decent standard of living for their families. Children of unmarried working mothers continue to be eligible for Medicaid and food stamps, and they benefit from the earned-income

tax credit that is available only to poorer working parents with children.

One of the most important, if hardest to document, gains from taking families off welfare is their greater self-respect when they provide for themselves. Mothers on welfare convey the impression to their children that it is normal to live off government handouts. In such an environment, it is difficult for children to place a high value on doing well at school and preparing for work by seeking out training on jobs and in schools.

Welfare reform has been a resounding success in inducing unmarried mothers to find jobs. This revolutionary approach to welfare is based on the appreciation that the vast majority of families do much better when treated as responsible adults and offered effective incentives to help themselves.

SOURCE: *Business Week*, May 24, 1999, p. 18.

person would be no more than four times the income of the poorest person. Although the measurement of inequality is difficult, it is clear that our society has much more inequality than Plato recommended.

One of the *Ten Principles of Economics* discussed in Chapter 1 is that governments can sometimes improve market outcomes. There is little consensus, however, about how this principle should be applied to the distribution of income. Philosophers and policymakers today do not agree on how much income inequality is desirable, or even whether public policy should aim to alter the distribution of income. Much of public debate reflects this disagreement. Whenever taxes are raised, for instance, lawmakers argue over how much of the tax hike should fall on the rich, the middle class, and the poor.

Another of the *Ten Principles of Economics* is that people face tradeoffs. This principle is important to keep in mind when thinking about economic inequality. Policies that penalize the successful and reward the unsuccessful reduce the incentive to succeed. Thus, policymakers face a tradeoff between equality and efficiency. The more equally the pie is divided, the smaller the pie becomes. This is the one lesson concerning the distribution of income about which almost everyone agrees.



Summary

- ◆ Data on the distribution of income show wide disparity in our society. The richest fifth of families earns about ten times as much income as the poorest fifth.
- ◆ Because in-kind transfers, the economic life cycle, transitory income, and economic mobility are so important for understanding variation in income, it is difficult to gauge the degree of inequality in our society using data on the distribution of income in a single year. When these factors are taken into account, they tend to suggest that economic well-being is more equally distributed than is annual income.
- ◆ Political philosophers differ in their views about the role of government in altering the distribution of income. Utilitarians (such as John Stuart Mill) would choose the distribution of income to maximize the sum of utility of everyone in society. Liberals (such as John Rawls) would determine the distribution of income as if we were behind a “veil of ignorance” that prevented us from knowing our own stations in life. Libertarians (such as Robert Nozick) would have the government enforce individual rights to ensure a fair process but then not be concerned about inequality in the resulting distribution of income.
- ◆ Various policies aim to help the poor—minimum-wage laws, welfare, negative income taxes, and in-kind transfers. Although each of these policies helps some families escape poverty, they also have unintended side effects. Because financial assistance declines as income rises, the poor often face effective marginal tax rates that are very high. Such high effective tax rates discourage poor families from escaping poverty on their own.

Key Concepts

poverty rate, p. 442
 poverty line, p. 442
 in-kind transfers, p. 444
 life cycle, p. 444

permanent income, p. 445
 utilitarianism, p. 447
 utility, p. 447
 liberalism, p. 448

maximin criterion, p. 448
 libertarianism, p. 450
 welfare, p. 452
 negative income tax, p. 454

Questions for Review

1. Does the richest fifth of the U.S. population earn two, four, or ten times the income of the poorest fifth?
2. How does the extent of income inequality in the United States compare to that of other nations around the world?
3. What groups in the population are most likely to live in poverty?
4. When gauging the amount of inequality, why do transitory and life cycle variations in income cause difficulties?
5. How would a utilitarian, a liberal, and a libertarian determine how much income inequality is permissible?
6. What are the pros and cons of in-kind (rather than cash) transfers to the poor?
7. Describe how antipoverty programs can discourage the poor from working. How might you reduce this disincentive? What are the disadvantages with your proposed policy?

Problems and Applications

1. Table 20-2 shows that income inequality in the United States has increased during the past 20 years. Some factors contributing to this increase were discussed in Chapter 19. What are they?
2. Table 20-4 shows that the percentage of children in families with income below the poverty line is almost twice the percentage of the elderly in such families. How might the allocation of government money across

different social programs have contributed to this phenomenon? (Hint: See Chapter 12.)

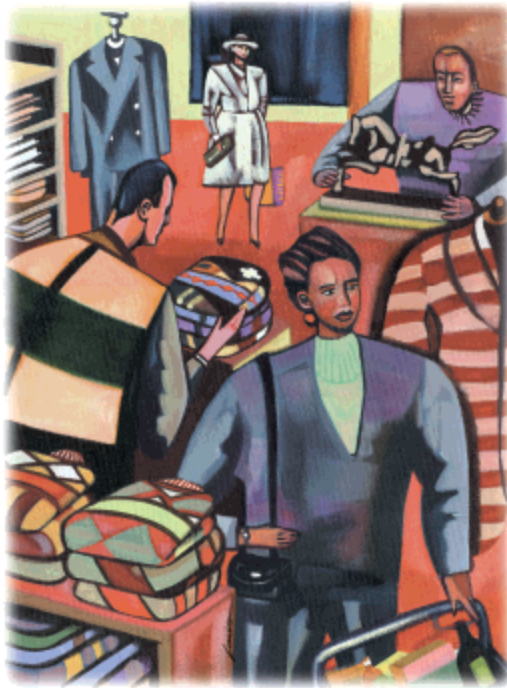
3. Economists often view life cycle variation in income as one form of transitory variation in income around people's lifetime, or permanent, income. In this sense, how does your current income compare to your permanent income? Do you think your current income accurately reflects your standard of living?
4. The chapter discusses the importance of economic mobility.
 - a. What policies might the government pursue to increase economic mobility *within* a generation?
 - b. What policies might the government pursue to increase economic mobility *across* generations?
 - c. Do you think we should reduce spending on current welfare programs in order to increase spending on programs that enhance economic mobility? What are some of the advantages and disadvantages of doing so?
5. Consider two communities. In one community, ten families have incomes of \$100 each and ten families have incomes of \$20 each. In the other community, ten families have incomes of \$200 each and ten families have incomes of \$22 each.
 - a. In which community is the distribution of income more unequal? In which community is the problem of poverty likely to be worse?
 - b. Which distribution of income would Rawls prefer? Explain.
 - c. Which distribution of income do you prefer? Explain.
6. The chapter uses the analogy of a "leaky bucket" to explain one constraint on the redistribution of income.
 - a. What elements of the U.S. system for redistributing income create the leaks in the bucket? Be specific.
 - b. Do you think that Republicans or Democrats generally believe that the bucket used for redistributing income is more leaky? How does that belief affect their views about the amount of income redistribution that the government should undertake?
7. Suppose there are two possible income distributions in a society of ten people. In the first distribution, nine people would have incomes of \$30,000 and one person would have an income of \$10,000. In the second distribution, all ten people would have incomes of \$25,000.
 - a. If the society had the first income distribution, what would be the utilitarian argument for redistributing income?
 - b. Which income distribution would Rawls consider more equitable? Explain.
 - c. Which income distribution would Nozick consider more equitable? Explain.
8. The poverty rate would be substantially lower if the market value of in-kind transfers were added to family income. The government spends more money on Medicaid than on any other in-kind transfer, with expenditures per recipient family amounting to roughly \$5,000 annually.
 - a. If the government gave each recipient family a check for this amount instead of enrolling them in the Medicaid program, do you think that most of these families would spend that much to purchase health insurance? (Recall that the poverty line is below \$15,000 for a family of four.) Why?
 - b. How does your answer to part (a) affect your view about whether we should determine the poverty rate by valuing in-kind transfers at the price the government pays for them? Explain.
 - c. How does your answer to part (a) affect your view about whether we should provide assistance to the poor in the form of cash transfers or in-kind transfers? Explain.
9. Suppose that a family's tax liability equaled its income multiplied by one-half, minus \$10,000. Under this system, some families would pay taxes to the government, and some families would receive money from the government through a "negative income tax."
 - a. Consider families with pre-tax incomes of \$0, \$10,000, \$20,000, \$30,000, and \$40,000. Make a table showing pre-tax income, taxes paid to the government or money received from the government, and after-tax income for each family.
 - b. What is the marginal tax rate in this system? (See Chapter 12 if you need to review the definition of marginal tax rate.) What is the maximum amount of income at which a family *receives* money from the government?
 - c. Now suppose that the tax schedule is changed, so that a family's tax liability equals its income multiplied by one-quarter, minus \$10,000. What is the marginal tax rate in this new system? What is the maximum amount of income at which a family receives money from the government?
 - d. What is the main advantage of each of the tax schedules discussed here?
10. John and Jeremy are utilitarians. John believes that labor supply is highly elastic, whereas Jeremy believes that

labor supply is quite inelastic. How do you suppose their views about income redistribution differ?

11. Do you agree or disagree with each of the following statements? What do your views imply for public policies, such as taxes on inheritance?

- a. "Every parent has the right to work hard and save in order to give his or her children a better life."
- b. "No child should be disadvantaged by the sloth or bad luck of his or her parents."

21



THE THEORY OF CONSUMER CHOICE

When you walk into a store, you are confronted with thousands of goods that you might buy. Of course, because your financial resources are limited, you cannot buy everything that you want. You therefore consider the prices of the various goods being offered for sale and buy a bundle of goods that, given your resources, best suits your needs and desires.

In this chapter we develop the theory that describes how consumers make decisions about what to buy. So far throughout this book, we have summarized consumers' decisions with the demand curve. As we discussed in Chapters 4 through 7, the demand curve for a good reflects consumers' willingness to pay for it. When the price of a good rises, consumers are willing to pay for fewer units, so the quantity demanded falls. We now look more deeply at the decisions that lie behind the demand curve. The theory of consumer choice presented in this chapter provides

IN THIS CHAPTER YOU WILL . . .

See how a budget constraint represents the choices a consumer can afford

Learn how indifference curves can be used to represent a consumer's preferences

Analyze how a consumer's optimal choices are determined

See how a consumer responds to changes in income and changes in prices

Decompose the impact of a price change into an income effect and a substitution effect

Apply the theory of consumer choice to four questions about household behavior



a more complete understanding of demand, just as the theory of the competitive firm in Chapter 14 provides a more complete understanding of supply.

One of the *Ten Principles of Economics* discussed in Chapter 1 is that people face tradeoffs. The theory of consumer choice examines the tradeoffs that people face in their role as consumers. When a consumer buys more of one good, he can afford less of other goods. When he spends more time enjoying leisure and less time working, he has lower income and can afford less consumption. When he spends more of his income in the present and saves less of it, he must accept a lower level of consumption in the future. The theory of consumer choice examines how consumers facing these tradeoffs make decisions and how they respond to changes in their environment.

After developing the basic theory of consumer choice, we apply it to several questions about household decisions. In particular, we ask:

- ◆ Do all demand curves slope downward?
- ◆ How do wages affect labor supply?
- ◆ How do interest rates affect household saving?
- ◆ Do the poor prefer to receive cash or in-kind transfers?

At first, these questions might seem unrelated. But, as we will see, we can use the theory of consumer choice to address each of them.

THE BUDGET CONSTRAINT: WHAT THE CONSUMER CAN AFFORD

Most people would like to increase the quantity or quality of the goods they consume—to take longer vacations, drive fancier cars, or eat at better restaurants. People consume less than they desire because their spending is *constrained*, or limited, by their income. We begin our study of consumer choice by examining this link between income and spending.

To keep things simple, we examine the decision facing a consumer who buys only two goods: Pepsi and pizza. Of course, real people buy thousands of different kinds of goods. Yet assuming there are only two goods greatly simplifies the problem without altering the basic insights about consumer choice.

We first consider how the consumer's income constrains the amount he spends on Pepsi and pizza. Suppose that the consumer has an income of \$1,000 per month and that he spends his entire income each month on Pepsi and pizza. The price of a pint of Pepsi is \$2, and the price of a pizza is \$10.

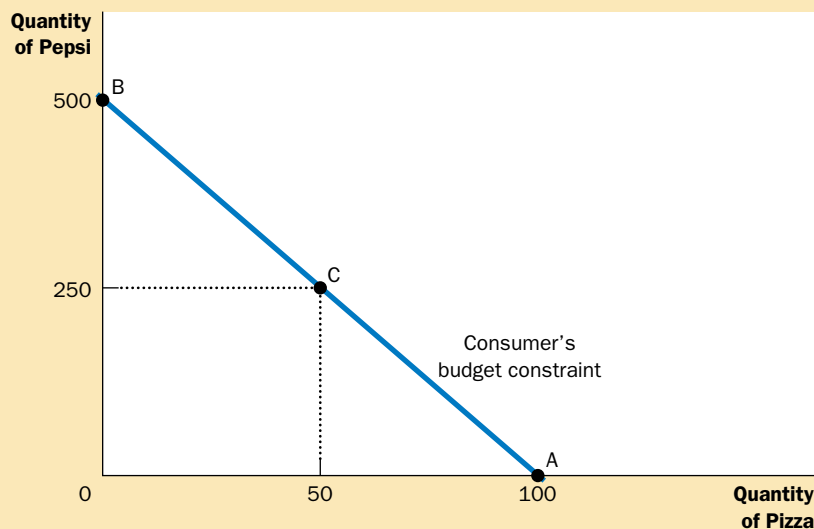
Table 21-1 shows some of the many combinations of Pepsi and pizza that the consumer can buy. The first line in the table shows that if the consumer spends all his income on pizza, he can eat 100 pizzas during the month, but he would not be able to buy any Pepsi at all. The second line shows another possible consumption bundle: 90 pizzas and 50 pints of Pepsi. And so on. Each consumption bundle in the table costs exactly \$1,000.

Figure 21-1 graphs the consumption bundles that the consumer can choose. The vertical axis measures the number of pints of Pepsi, and the horizontal axis

Table 21-1

PINTS OF PEPSI	NUMBER OF PIZZAS	SPENDING ON PEPSI	SPENDING ON PIZZA	TOTAL SPENDING
0	100	\$ 0	\$1,000	\$1,000
50	90	100	900	1,000
100	80	200	800	1,000
150	70	300	700	1,000
200	60	400	600	1,000
250	50	500	500	1,000
300	40	600	400	1,000
350	30	700	300	1,000
400	20	800	200	1,000
450	10	900	100	1,000
500	0	1,000	0	1,000

THE CONSUMER'S OPPORTUNITIES. This table shows what the consumer can afford if his income is \$1,000, the price of Pepsi is \$2, and the price of pizza is \$10.

Figure 21-1

THE CONSUMER'S BUDGET CONSTRAINT. The budget constraint shows the various bundles of goods that the consumer can afford for a given income. Here the consumer buys bundles of Pepsi and pizza. The more Pepsi he buys, the less pizza he can afford.

measures the number of pizzas. Three points are marked on this figure. At point A, the consumer buys no Pepsi and consumes 100 pizzas. At point B, the consumer buys no pizza and consumes 500 pints of Pepsi. At point C, the consumer buys 50 pizzas and 250 pints of Pepsi. Point C, which is exactly at the middle of the line from A to B, is the point at which the consumer spends an equal amount (\$500) on Pepsi and pizza. Of course, these are only three of the many combinations of Pepsi and pizza that the consumer can choose. All the points on the line from A to B are possible. This line, called the **budget constraint**, shows the consumption bundles that the consumer can afford. In this case, it shows the tradeoff between Pepsi and pizza that the consumer faces.

budget constraint
the limit on the consumption bundles that a consumer can afford

The slope of the budget constraint measures the rate at which the consumer can trade one good for the other. Recall from the appendix to Chapter 2 that the slope between two points is calculated as the change in the vertical distance divided by the change in the horizontal distance (“rise over run”). From point A to point B, the vertical distance is 500 pints, and the horizontal distance is 100 pizzas. Thus, the slope is 5 pints per pizza. (Actually, because the budget constraint slopes downward, the slope is a negative number. But for our purposes we can ignore the minus sign.)

Notice that the slope of the budget constraint equals the *relative price* of the two goods—the price of one good compared to the price of the other. A pizza costs 5 times as much as a pint of Pepsi, so the opportunity cost of a pizza is 5 pints of Pepsi. The budget constraint’s slope of 5 reflects the tradeoff the market is offering the consumer: 1 pizza for 5 pints of Pepsi.

QUICK QUIZ: Draw the budget constraint for a person with income of \$1,000 if the price of Pepsi is \$5 and the price of pizza is \$10. What is the slope of this budget constraint?

PREFERENCES: WHAT THE CONSUMER WANTS

Our goal in this chapter is to see how consumers make choices. The budget constraint is one piece of the analysis: It shows what combination of goods the consumer can afford given his income and the prices of the goods. The consumer’s choices, however, depend not only on his budget constraint but also on his preferences regarding the two goods. Therefore, the consumer’s preferences are the next piece of our analysis.

REPRESENTING PREFERENCES WITH INDIFFERENCE CURVES

The consumer’s preferences allow him to choose among different bundles of Pepsi and pizza. If you offer the consumer two different bundles, he chooses the bundle that best suits his tastes. If the two bundles suit his tastes equally well, we say that the consumer is *indifferent* between the two bundles.

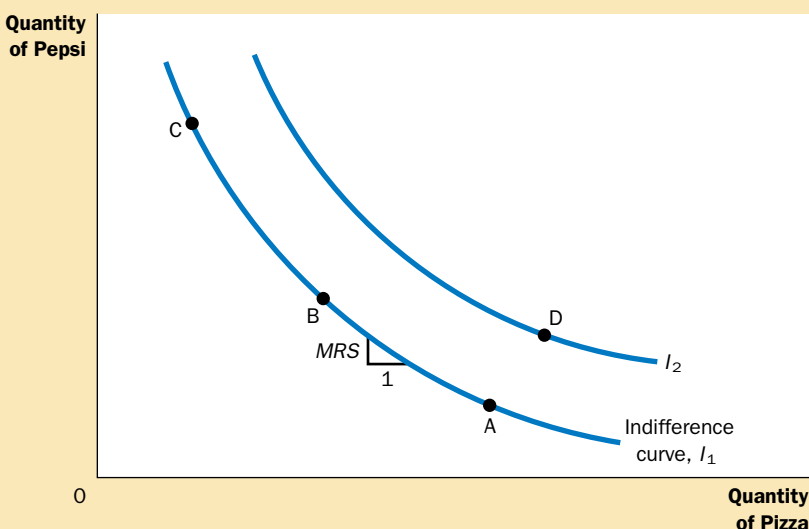
Just as we have represented the consumer’s budget constraint graphically, we can also represent his preferences graphically. We do this with indifference curves. An **indifference curve** shows the bundles of consumption that make the consumer equally happy. In this case, the indifference curves show the combinations of Pepsi and pizza with which the consumer is equally satisfied.

Figure 21-2 shows two of the consumer’s many indifference curves. The consumer is indifferent among combinations A, B, and C, because they are all on the same curve. Not surprisingly, if the consumer’s consumption of pizza is reduced, say from point A to point B, consumption of Pepsi must increase to keep him equally happy. If consumption of pizza is reduced again, from point B to point C, the amount of Pepsi consumed must increase yet again.

indifference curve

a curve that shows consumption bundles that give the consumer the same level of satisfaction

Figure 21-2



THE CONSUMER'S PREFERENCES. The consumer's preferences are represented with indifference curves, which show the combinations of Pepsi and pizza that make the consumer equally satisfied. Because the consumer prefers more of a good, points on a higher indifference curve (I_2 here) are preferred to points on a lower indifference curve (I_1). The marginal rate of substitution (MRS) shows the rate at which the consumer is willing to trade Pepsi for pizza.

The slope at any point on an indifference curve equals the rate at which the consumer is willing to substitute one good for the other. This rate is called the **marginal rate of substitution (MRS)**. In this case, the marginal rate of substitution measures how much Pepsi the consumer requires in order to be compensated for a one-unit reduction in pizza consumption. Notice that because the indifference curves are not straight lines, the marginal rate of substitution is not the same at all points on a given indifference curve. The rate at which a consumer is willing to trade one good for the other depends on the amounts of the goods he is already consuming. That is, the rate at which a consumer is willing to trade pizza for Pepsi depends on whether he is more hungry or more thirsty, which in turn depends on how much pizza and Pepsi he has.

The consumer is equally happy at all points on any given indifference curve, but he prefers some indifference curves to others. Because he prefers more consumption to less, higher indifference curves are preferred to lower ones. In Figure 21-2, any point on curve I_2 is preferred to any point on curve I_1 .

A consumer's set of indifference curves gives a complete ranking of the consumer's preferences. That is, we can use the indifference curves to rank any two bundles of goods. For example, the indifference curves tell us that point D is preferred to point A because point D is on a higher indifference curve than point A. (That conclusion may be obvious, however, because point D offers the consumer both more pizza and more Pepsi.) The indifference curves also tell us that point D is preferred to point C because point D is on a higher indifference curve. Even though point D has less Pepsi than point C, it has more than enough extra pizza to make the consumer prefer it. By seeing which point is on the higher indifference curve, we can use the set of indifference curves to rank any combinations of Pepsi and pizza.

marginal rate of substitution

the rate at which a consumer is willing to trade one good for another

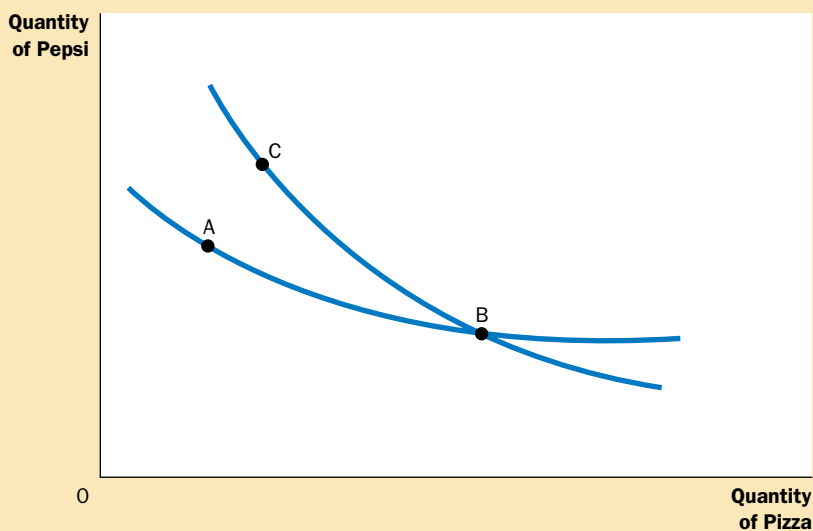
FOUR PROPERTIES OF INDIFFERENCE CURVES

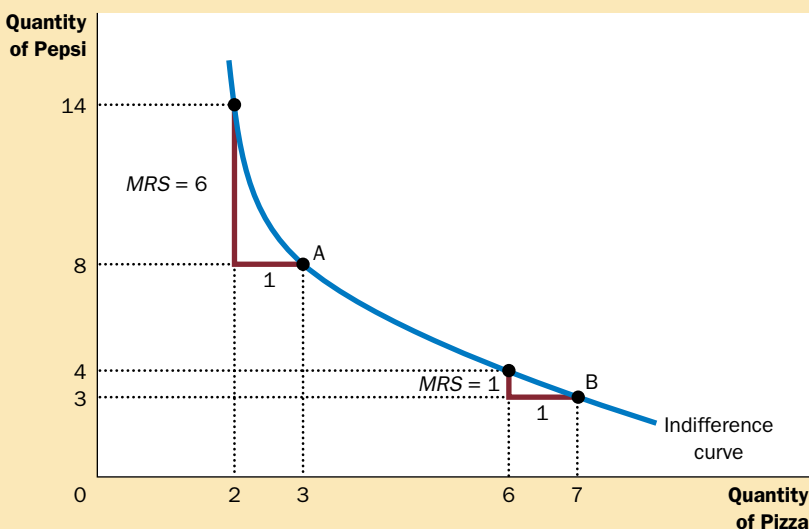
Because indifference curves represent a consumer's preferences, they have certain properties that reflect those preferences. Here we consider four properties that describe most indifference curves:

- ◆ *Property 1: Higher indifference curves are preferred to lower ones.* Consumers usually prefer more of something to less of it. (That is why we call this something a “good” rather than a “bad.”) This preference for greater quantities is reflected in the indifference curves. As Figure 21-2 shows, higher indifference curves represent larger quantities of goods than lower indifference curves. Thus, the consumer prefers being on higher indifference curves.
- ◆ *Property 2: Indifference curves are downward sloping.* The slope of an indifference curve reflects the rate at which the consumer is willing to substitute one good for the other. In most cases, the consumer likes both goods. Therefore, if the quantity of one good is reduced, the quantity of the other good must increase in order for the consumer to be equally happy. For this reason, most indifference curves slope downward.
- ◆ *Property 3: Indifference curves do not cross.* To see why this is true, suppose that two indifference curves did cross, as in Figure 21-3. Then, because point A is on the same indifference curve as point B, the two points would make the consumer equally happy. In addition, because point B is on the same indifference curve as point C, these two points would make the consumer equally happy. But these conclusions imply that points A and C would also make the consumer equally happy, even though point C has more of both goods. This contradicts our assumption that the consumer always prefers more of both goods to less. Thus, indifference curves cannot cross.

Figure 21-3

THE IMPOSSIBILITY OF INTERSECTING INDIFFERENCE CURVES. A situation like this can never happen. According to these indifference curves, the consumer would be equally satisfied at points A, B, and C, even though point C has more of both goods than point A.



**Figure 21-4**

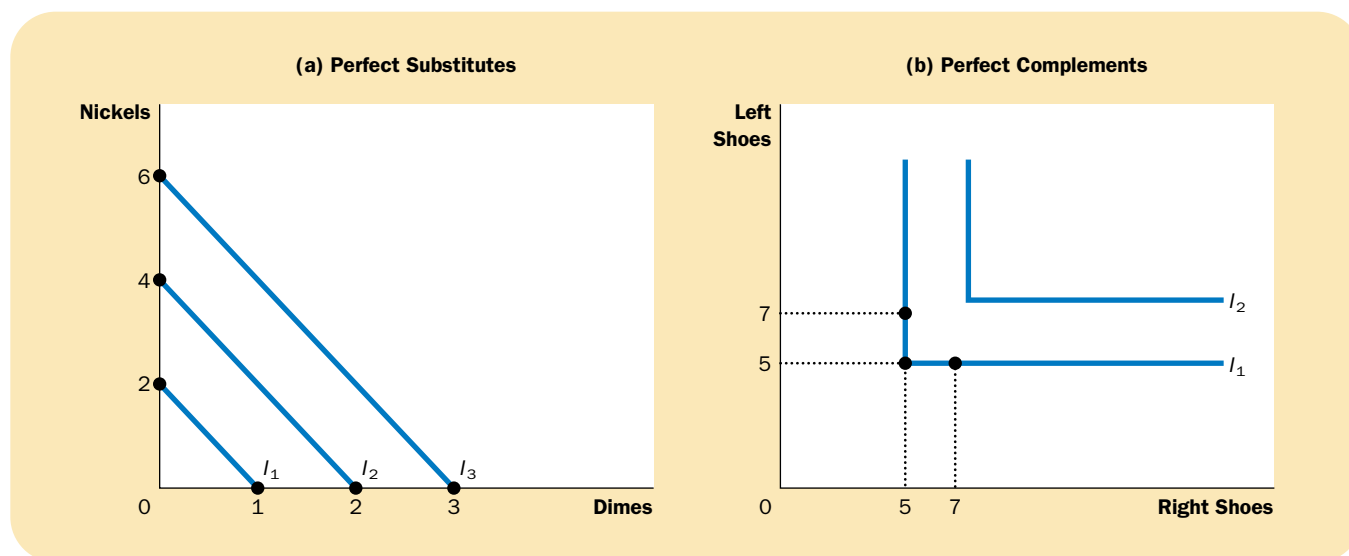
BOWED INDIFFERENCE CURVES. Indifference curves are usually bowed inward. This shape implies that the marginal rate of substitution (MRS) depends on the quantity of the two goods the consumer is consuming. At point A, the consumer has little pizza and much Pepsi, so he requires a lot of extra Pepsi to induce him to give up one of the pizzas: The marginal rate of substitution is 6 pints of Pepsi per pizza. At point B, the consumer has much pizza and little Pepsi, so he requires only a little extra Pepsi to induce him to give up one of the pizzas: The marginal rate of substitution is 1 pint of Pepsi per pizza.

- ◆ **Property 4: Indifference curves are bowed inward.** The slope of an indifference curve is the marginal rate of substitution—the rate at which the consumer is willing to trade off one good for the other. The marginal rate of substitution (MRS) usually depends on the amount of each good the consumer is currently consuming. In particular, because people are more willing to trade away goods that they have in abundance and less willing to trade away goods of which they have little, the indifference curves are bowed inward. As an example, consider Figure 21-4. At point A, because the consumer has a lot of Pepsi and only a little pizza, he is very hungry but not very thirsty. To induce the consumer to give up 1 pizza, the consumer has to be given 6 pints of Pepsi: The marginal rate of substitution is 6 pints per pizza. By contrast, at point B, the consumer has little Pepsi and a lot of pizza, so he is very thirsty but not very hungry. At this point, he would be willing to give up 1 pizza to get 1 pint of Pepsi: The marginal rate of substitution is 1 pint per pizza. Thus, the bowed shape of the indifference curve reflects the consumer's greater willingness to give up a good that he already has in large quantity.

TWO EXTREME EXAMPLES OF INDIFFERENCE CURVES

The shape of an indifference curve tells us about the consumer's willingness to trade one good for the other. When the goods are easy to substitute for each other, the indifference curves are less bowed; when the goods are hard to substitute, the indifference curves are very bowed. To see why this is true, let's consider the extreme cases.

Perfect Substitutes Suppose that someone offered you bundles of nickels and dimes. How would you rank the different bundles?

**Figure 21-5**

PERFECT SUBSTITUTES AND PERFECT COMPLEMENTS. When two goods are easily substitutable, such as nickels and dimes, the indifference curves are straight lines, as shown in panel (a). When two goods are strongly complementary, such as left shoes and right shoes, the indifference curves are right angles, as shown in panel (b).

Most likely, you would care only about the total monetary value of each bundle. If so, you would judge a bundle based on the number of nickels plus twice the number of dimes. In other words, you would always be willing to trade 1 dime for 2 nickels, regardless of the number of nickels and dimes in the bundle. Your marginal rate of substitution between nickels and dimes would be a fixed number—2.

We can represent your preferences over nickels and dimes with the indifference curves in panel (a) of Figure 21-5. Because the marginal rate of substitution is constant, the indifference curves are straight lines. In this extreme case of straight indifference curves, we say that the two goods are **perfect substitutes**.

perfect substitutes
two goods with straight-line
indifference curves

Perfect Complements Suppose now that someone offered you bundles of shoes. Some of the shoes fit your left foot, others your right foot. How would you rank these different bundles?

In this case, you might care only about the number of pairs of shoes. In other words, you would judge a bundle based on the number of pairs you could assemble from it. A bundle of 5 left shoes and 7 right shoes yields only 5 pairs. Getting 1 more right shoe has no value if there is no left shoe to go with it.

We can represent your preferences for right and left shoes with the indifference curves in panel (b) of Figure 21-5. In this case, a bundle with 5 left shoes and 5 right shoes is just as good as a bundle with 5 left shoes and 7 right shoes. It is also just as good as a bundle with 7 left shoes and 5 right shoes. The indifference curves, therefore, are right angles. In this extreme case of right-angle indifference curves, we say that the two goods are **perfect complements**.

perfect complements
two goods with right-angle
indifference curves

In the real world, of course, most goods are neither perfect substitutes (like nickels and dimes) nor perfect complements (like right shoes and left shoes). More

FYI

Utility: An Alternative Way to Represent a Consumer's Preferences



We have used indifference curves to represent the consumer's preferences. Another common way to represent preferences is with the concept of *utility*. Utility is an abstract measure of the satisfaction or happiness that a consumer receives from a bundle of goods. Economists say that a consumer prefers one bundle of goods to another if the first provides more utility than the second.

Indifference curves and utility are closely related. Because the consumer prefers points on higher indifference

curves, bundles of goods on higher indifference curves provide higher utility. Because the consumer is equally happy with all points on the same indifference curve, all these bundles provide the same utility. Indeed, you can think of an indifference curve as an "equal-utility" curve. The slope of the indifference curve (the marginal rate of substitution) reflects the marginal utility generated by one good compared to the marginal utility generated by the other good.

When economists discuss the theory of consumer choice, they might express the theory using different words. One economist might say that the goal of the consumer is to maximize utility. Another might say that the goal of the consumer is to end up on the highest possible indifference curve. In essence, these are two ways of saying the same thing.

typically, the indifference curves are bowed inward, but not so bowed as to become right angles.

QUICK QUIZ: Draw some indifference curves for Pepsi and pizza. Explain the four properties of these indifference curves.

OPTIMIZATION: WHAT THE CONSUMER CHOOSES

The goal of this chapter is to understand how a consumer makes choices. We have the two pieces necessary for this analysis: the consumer's budget constraint and the consumer's preferences. Now we put these two pieces together and consider the consumer's decision about what to buy.

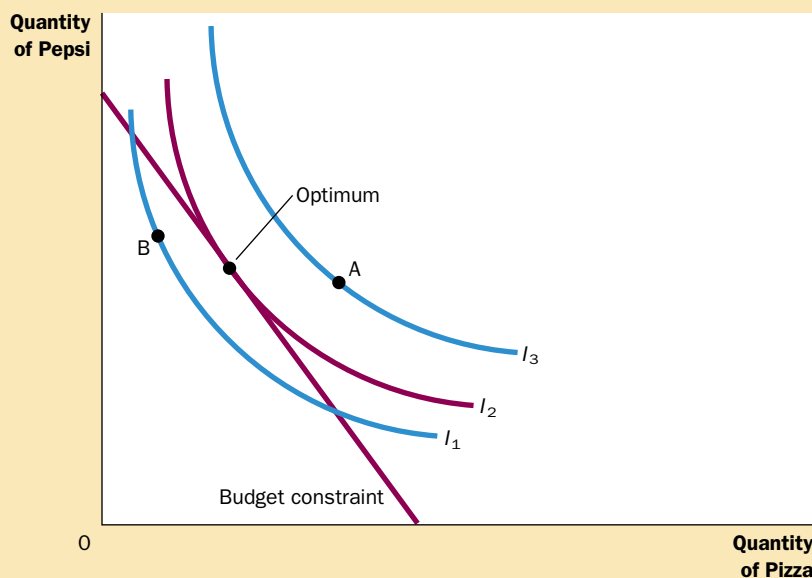
THE CONSUMER'S OPTIMAL CHOICES

Consider once again our Pepsi and pizza example. The consumer would like to end up with the best possible combination of Pepsi and pizza—that is, the combination on the highest possible indifference curve. But the consumer must also end up on or below his budget constraint, which measures the total resources available to him.

Figure 21-6 shows the consumer's budget constraint and three of his many indifference curves. The highest indifference curve that the consumer can reach (I_2 in the figure) is the one that just barely touches the budget constraint. The point at which this indifference curve and the budget constraint touch is called the *optimum*. The consumer would prefer point A, but he cannot afford that point because

Figure 21-6**THE CONSUMER'S OPTIMUM.**

The consumer chooses the point on his budget constraint that lies on the highest indifference curve. At this point, called the optimum, the marginal rate of substitution equals the relative price of the two goods. Here the highest indifference curve the consumer can reach is I_2 . The consumer prefers point A, which lies on indifference curve I_3 , but the consumer cannot afford this bundle of Pepsi and pizza. By contrast, point B is affordable, but because it lies on a lower indifference curve, the consumer does not prefer it.



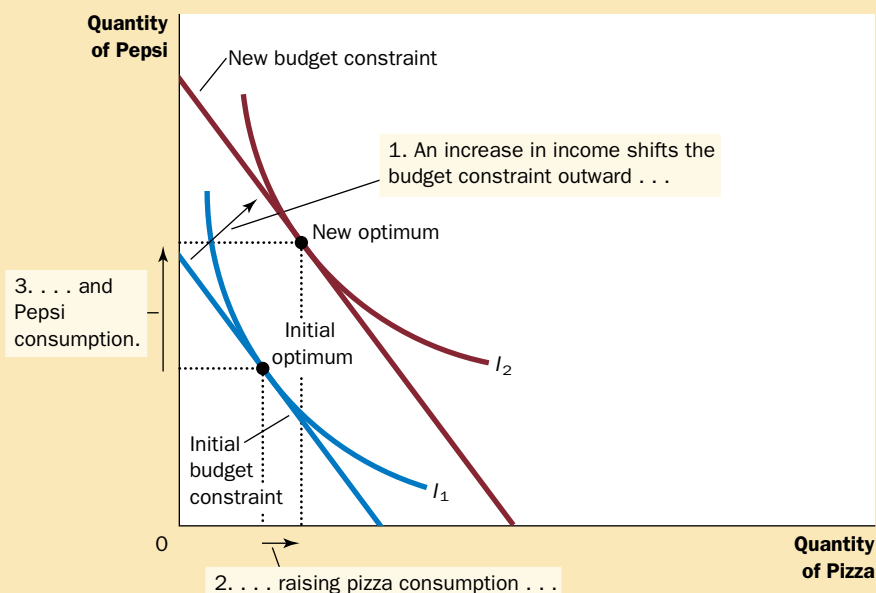
it lies above his budget constraint. The consumer can afford point B, but that point is on a lower indifference curve and, therefore, provides the consumer less satisfaction. The optimum represents the best combination of consumption of Pepsi and pizza available to the consumer.

Notice that, at the optimum, the slope of the indifference curve equals the slope of the budget constraint. We say that the indifference curve is *tangent* to the budget constraint. The slope of the indifference curve is the marginal rate of substitution between Pepsi and pizza, and the slope of the budget constraint is the relative price of Pepsi and pizza. Thus, *the consumer chooses consumption of the two goods so that the marginal rate of substitution equals the relative price.*

In Chapter 7 we saw how market prices reflect the marginal value that consumers place on goods. This analysis of consumer choice shows the same result in another way. In making his consumption choices, the consumer takes as given the relative price of the two goods and then chooses an optimum at which his marginal rate of substitution equals this relative price. The relative price is the rate at which the *market* is willing to trade one good for the other, whereas the marginal rate of substitution is the rate at which the *consumer* is willing to trade one good for the other. At the consumer's optimum, the consumer's valuation of the two goods (as measured by the marginal rate of substitution) equals the market's valuation (as measured by the relative price). As a result of this consumer optimization, market prices of different goods reflect the value that consumers place on those goods.

HOW CHANGES IN INCOME AFFECT THE CONSUMER'S CHOICES

Now that we have seen how the consumer makes the consumption decision, let's examine how consumption responds to changes in income. To be specific, suppose

Figure 21-7

AN INCREASE IN INCOME. When the consumer's income rises, the budget constraint shifts out. If both goods are normal goods, the consumer responds to the increase in income by buying more of both of them. Here the consumer buys more pizza and more Pepsi.

that income increases. With higher income, the consumer can afford more of both goods. The increase in income, therefore, shifts the budget constraint outward, as in Figure 21-7. Because the relative price of the two goods has not changed, the slope of the new budget constraint is the same as the slope of the initial budget constraint. That is, an increase in income leads to a parallel shift in the budget constraint.

The expanded budget constraint allows the consumer to choose a better combination of Pepsi and pizza. In other words, the consumer can now reach a higher indifference curve. Given the shift in the budget constraint and the consumer's preferences as represented by his indifference curves, the consumer's optimum moves from the point labeled "initial optimum" to the point labeled "new optimum."

Notice that, in Figure 21-7, the consumer chooses to consume more Pepsi and more pizza. Although the logic of the model does not require increased consumption of both goods in response to increased income, this situation is the most common one. As you may recall from Chapter 4, if a consumer wants more of a good when his income rises, economists call it a **normal good**. The indifference curves in Figure 21-7 are drawn under the assumption that both Pepsi and pizza are normal goods.

Figure 21-8 shows an example in which an increase in income induces the consumer to buy more pizza but less Pepsi. If a consumer buys less of a good when his income rises, economists call it an **inferior good**. Figure 21-8 is drawn under the assumption that pizza is a normal good and Pepsi is an inferior good.

Although most goods are normal goods, there are some inferior goods in the world. One example is bus rides. High-income consumers are more likely to own cars and less likely to ride the bus than low-income consumers. Bus rides, therefore, are an inferior good.

normal good

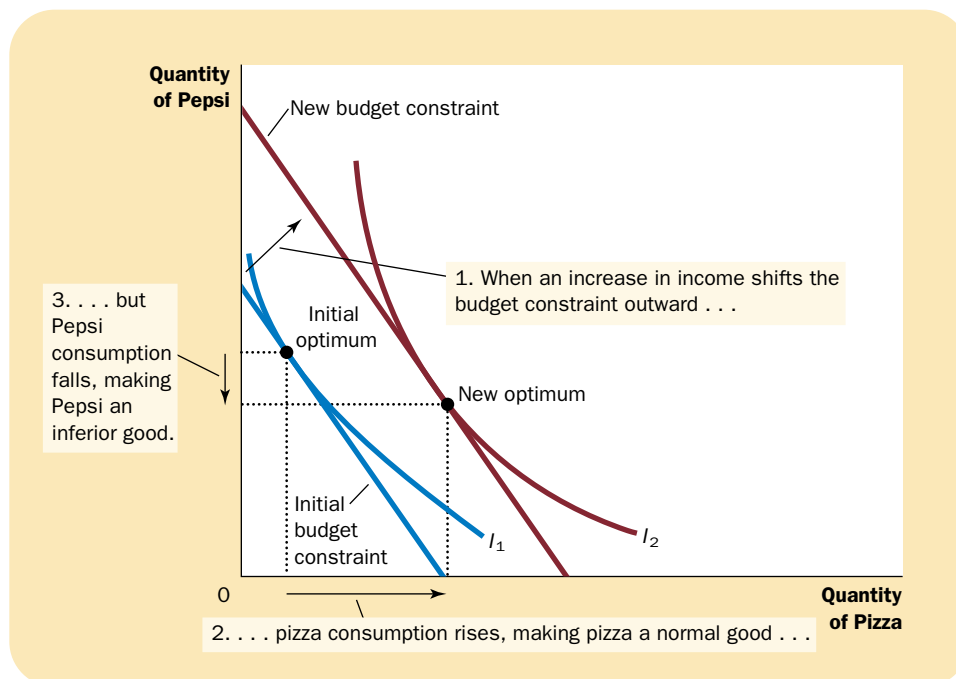
a good for which an increase in income raises the quantity demanded

inferior good

a good for which an increase in income reduces the quantity demanded

Figure 21-8

AN INFERIOR GOOD. A good is an inferior good if the consumer buys less of it when his income rises. Here Pepsi is an inferior good: When the consumer's income increases and the budget constraint shifts outward, the consumer buys more pizza but less Pepsi.



HOW CHANGES IN PRICES AFFECT THE CONSUMER'S CHOICES

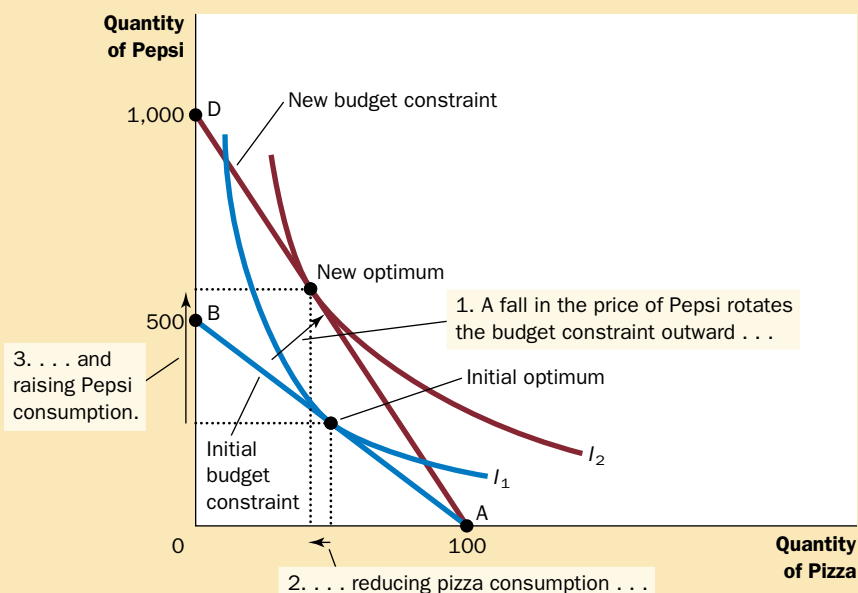
Let's now use this model of consumer choice to consider how a change in the price of one of the goods alters the consumer's choices. Suppose, in particular, that the price of Pepsi falls from \$2 to \$1 a pint. It is no surprise that the lower price expands the consumer's set of buying opportunities. In other words, a fall in the price of any good shifts the budget constraint outward.

Figure 21-9 considers more specifically how the fall in price affects the budget constraint. If the consumer spends his entire \$1,000 income on pizza, then the price of Pepsi is irrelevant. Thus, point A in the figure stays the same. Yet if the consumer spends his entire income of \$1,000 on Pepsi, he can now buy 1,000 rather than only 500 pints. Thus, the end point of the budget constraint moves from point B to point D.

Notice that in this case the outward shift in the budget constraint changes its slope. (This differs from what happened previously when prices stayed the same but the consumer's income changed.) As we have discussed, the slope of the budget constraint reflects the relative price of Pepsi and pizza. Because the price of Pepsi has fallen to \$1 from \$2, while the price of pizza has remained \$10, the consumer can now trade a pizza for 10 rather than 5 pints of Pepsi. As a result, the new budget constraint is more steeply sloped.

How such a change in the budget constraint alters the consumption of both goods depends on the consumer's preferences. For the indifference curves drawn in this figure, the consumer buys more Pepsi and less pizza.

Figure 21-9



A CHANGE IN PRICE. When the price of Pepsi falls, the consumer's budget constraint shifts outward and changes slope. The consumer moves from the initial optimum to the new optimum, which changes his purchases of both Pepsi and pizza. In this case, the quantity of Pepsi consumed rises, and the quantity of pizza consumed falls.

INCOME AND SUBSTITUTION EFFECTS

The impact of a change in the price of a good on consumption can be decomposed into two effects: an **income effect** and a **substitution effect**. To see what these two effects are, consider how our consumer might respond when he learns that the price of Pepsi has fallen. He might reason in the following ways:

- ◆ "Great news! Now that Pepsi is cheaper, my income has greater purchasing power. I am, in effect, richer than I was. Because I am richer, I can buy both more Pepsi and more pizza." (This is the income effect.)
- ◆ "Now that the price of Pepsi has fallen, I get more pints of Pepsi for every pizza that I give up. Because pizza is now relatively more expensive, I should buy less pizza and more Pepsi." (This is the substitution effect.)

Which statement do you find more compelling?

In fact, both of these statements make sense. The decrease in the price of Pepsi makes the consumer better off. If Pepsi and pizza are both normal goods, the consumer will want to spread this improvement in his purchasing power over both goods. This income effect tends to make the consumer buy more pizza and more Pepsi. Yet, at the same time, consumption of Pepsi has become less expensive relative to consumption of pizza. This substitution effect tends to make the consumer choose more Pepsi and less pizza.

Now consider the end result of these two effects. The consumer certainly buys more Pepsi, because the income and substitution effects both act to raise purchases of Pepsi. But it is ambiguous whether the consumer buys more pizza, because the

income effect

the change in consumption that results when a price change moves the consumer to a higher or lower indifference curve

substitution effect

the change in consumption that results when a price change moves the consumer along a given indifference curve to a point with a new marginal rate of substitution

income and substitution effects work in opposite directions. This conclusion is summarized in Table 21-2.

We can interpret the income and substitution effects using indifference curves. *The income effect is the change in consumption that results from the movement to a higher indifference curve. The substitution effect is the change in consumption that results from being at a point on an indifference curve with a different marginal rate of substitution.*

Figure 21-10 shows graphically how to decompose the change in the consumer's decision into the income effect and the substitution effect. When the price

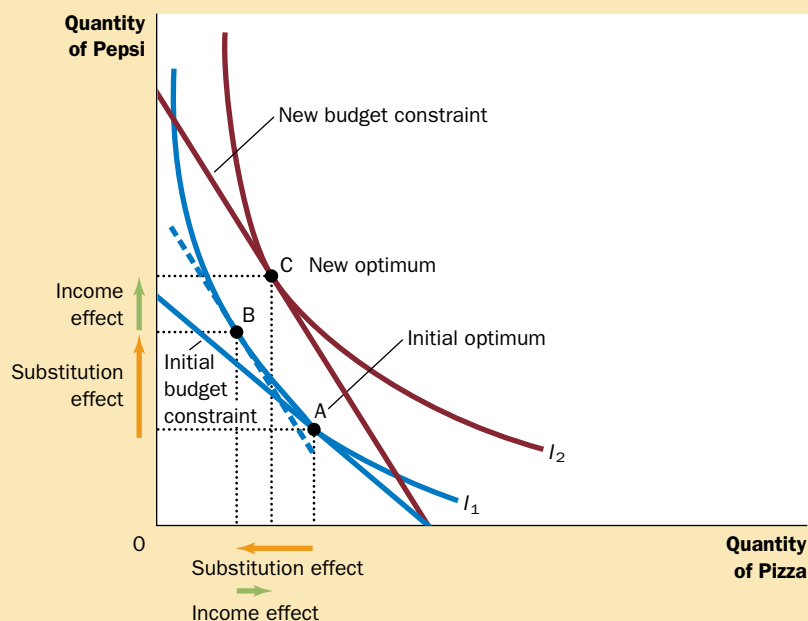
GOOD	INCOME EFFECT	SUBSTITUTION EFFECT	TOTAL EFFECT
Pepsi	Consumer is richer, so he buys more Pepsi.	Pepsi is relatively cheaper, so consumer buys more Pepsi.	Income and substitution effects act in same direction, so consumer buys more Pepsi.
Pizza	Consumer is richer, so he buys more pizza.	Pizza is relatively more expensive, so consumer buys less pizza.	Income and substitution effects act in opposite directions, so the total effect on pizza consumption is ambiguous.

Table 21-2

INCOME AND SUBSTITUTION EFFECTS WHEN THE PRICE OF PEPSI FALLS

Figure 21-10

INCOME AND SUBSTITUTION EFFECTS. The effect of a change in price can be broken down into an income effect and a substitution effect. The substitution effect—the movement along an indifference curve to a point with a different marginal rate of substitution—is shown here as the change from point A to point B along indifference curve I_1 . The income effect—the shift to a higher indifference curve—is shown here as the change from point B on indifference curve I_1 to point C on indifference curve I_2 .



of Pepsi falls, the consumer moves from the initial optimum, point A, to the new optimum, point C. We can view this change as occurring in two steps. First, the consumer moves *along* the initial indifference curve I_1 from point A to point B. The consumer is equally happy at these two points, but at point B, the marginal rate of substitution reflects the new relative price. (The dashed line through point B reflects the new relative price by being parallel to the new budget constraint.) Next, the consumer *shifts* to the higher indifference curve I_2 by moving from point B to point C. Even though point B and point C are on different indifference curves, they have the same marginal rate of substitution. That is, the slope of the indifference curve I_1 at point B equals the slope of the indifference curve I_2 at point C.

Although the consumer never actually chooses point B, this hypothetical point is useful to clarify the two effects that determine the consumer's decision. Notice that the change from point A to point B represents a pure change in the marginal rate of substitution without any change in the consumer's welfare. Similarly, the change from point B to point C represents a pure change in welfare without any change in the marginal rate of substitution. Thus, the movement from A to B shows the substitution effect, and the movement from B to C shows the income effect.

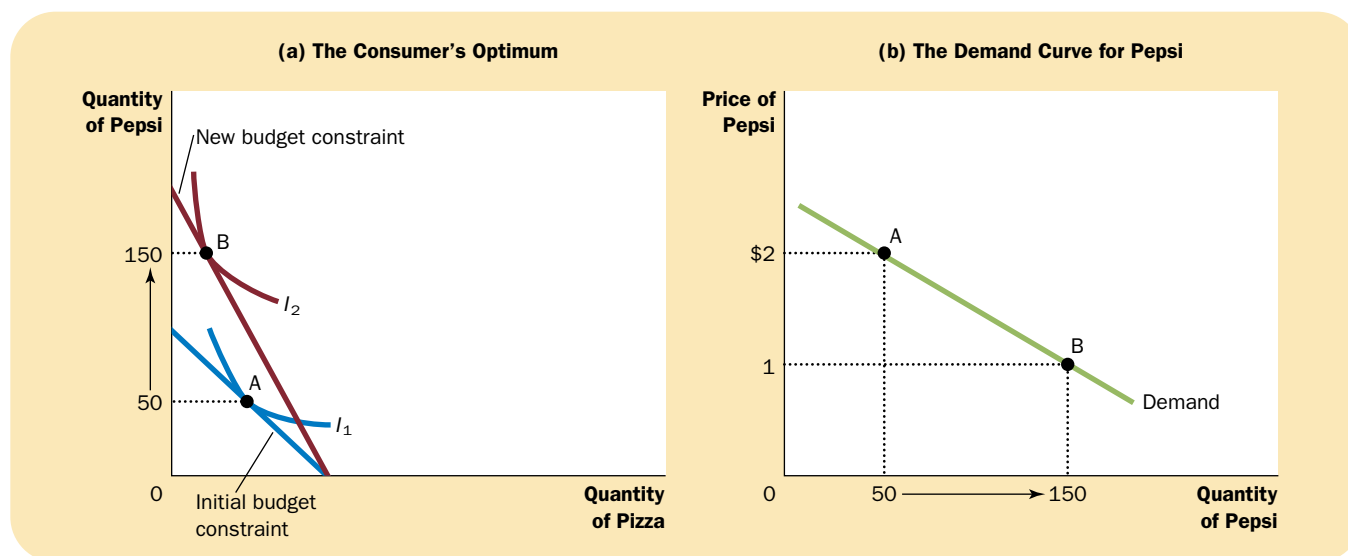
DERIVING THE DEMAND CURVE

We have just seen how changes in the price of a good alter the consumer's budget constraint and, therefore, the quantities of the two goods that he chooses to buy. The demand curve for any good reflects these consumption decisions. Recall that a demand curve shows the quantity demanded of a good for any given price. We can view a consumer's demand curve as a summary of the optimal decisions that arise from his budget constraint and indifference curves.

For example, Figure 21-11 considers the demand for Pepsi. Panel (a) shows that when the price of a pint falls from \$2 to \$1, the consumer's budget constraint shifts outward. Because of both income and substitution effects, the consumer increases his purchases of Pepsi from 50 to 150 pints. Panel (b) shows the demand curve that results from this consumer's decisions. In this way, the theory of consumer choice provides the theoretical foundation for the consumer's demand curve, which we first introduced in Chapter 4.

Although it is comforting to know that the demand curve arises naturally from the theory of consumer choice, this exercise by itself does not justify developing the theory. There is no need for a rigorous, analytic framework just to establish that people respond to changes in prices. The theory of consumer choice is, however, very useful. As we see in the next section, we can use the theory to delve more deeply into the determinants of household behavior.

QUICK QUIZ: Draw a budget constraint and indifference curves for Pepsi and pizza. Show what happens to the budget constraint and the consumer's optimum when the price of pizza rises. In your diagram, decompose the change into an income effect and a substitution effect.

**Figure 21-11**

DERIVING THE DEMAND CURVE. Panel (a) shows that when the price of Pepsi falls from \$2 to \$1, the consumer's optimum moves from point A to point B, and the quantity of Pepsi consumed rises from 50 to 150 pints. The demand curve in panel (b) reflects this relationship between the price and the quantity demanded.

FOUR APPLICATIONS

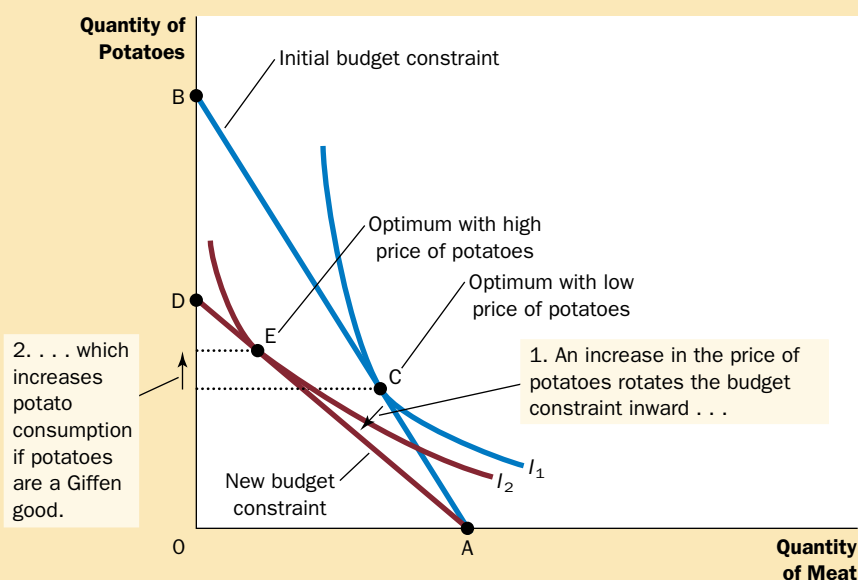
Now that we have developed the basic theory of consumer choice, let's use it to shed light on four questions about how the economy works. These four questions might at first seem unrelated. But because each question involves household decisionmaking, we can address it with the model of consumer behavior we have just developed.

DO ALL DEMAND CURVES SLOPE DOWNWARD?

Normally, when the price of a good rises, people buy less of it. Chapter 4 called this usual behavior the *law of demand*. This law is reflected in the downward slope of the demand curve.

As a matter of economic theory, however, demand curves can sometimes slope upward. In other words, consumers can sometimes violate the law of demand and buy *more* of a good when the price rises. To see how this can happen, consider Figure 21-12. In this example, the consumer buys two goods—meat and potatoes. Initially, the consumer's budget constraint is the line from point A to point B. The optimum is point C. When the price of potatoes rises, the budget constraint shifts inward and is now the line from point A to point D. The optimum is now point E.

Figure 21-12



A GIFFEN GOOD. In this example, when the price of potatoes rises, the consumer's optimum shifts from point C to point E. In this case, the consumer responds to a higher price of potatoes by buying less meat and more potatoes.

Notice that a rise in the price of potatoes has led the consumer to buy a larger quantity of potatoes.

Why is the consumer responding in a seemingly perverse way? The reason is that potatoes here are a strongly inferior good. When the price of potatoes rises, the consumer is poorer. The income effect makes the consumer want to buy less meat and more potatoes. At the same time, because the potatoes have become more expensive relative to meat, the substitution effect makes the consumer want to buy more meat and less potatoes. In this particular case, however, the income effect is so strong that it exceeds the substitution effect. In the end, the consumer responds to the higher price of potatoes by buying less meat and more potatoes.

Economists use the term **Giffen good** to describe a good that violates the law of demand. (The term is named for economist Robert Giffen, who first noted this possibility.) In this example, potatoes are a Giffen good. Giffen goods are inferior goods for which the income effect dominates the substitution effect. Therefore, they have demand curves that slope upward.

Economists disagree about whether any Giffen good has ever been discovered. Some historians suggest that potatoes were in fact a Giffen good during the Irish potato famine of the nineteenth century. Potatoes were such a large part of people's diet that when the price of potatoes rose, it had a large income effect. People responded to their reduced living standard by cutting back on the luxury of meat and buying more of the staple food of potatoes. Thus, it is argued that a higher price of potatoes actually raised the quantity of potatoes demanded.

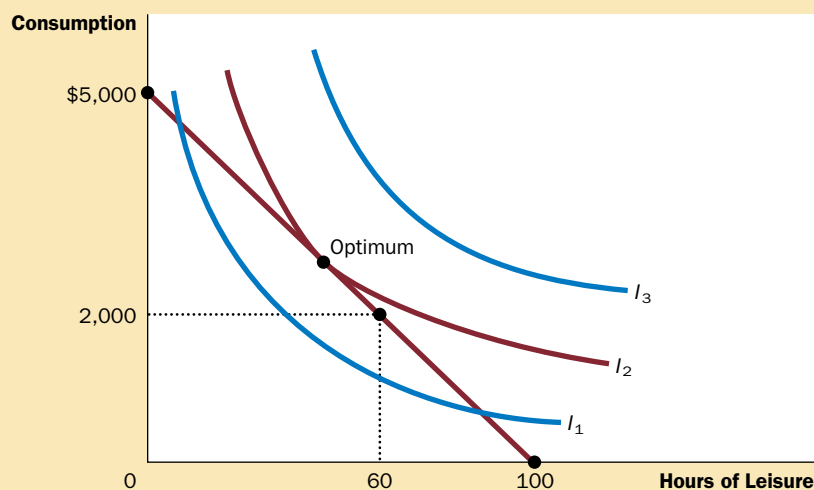
Whether or not this historical account is true, it is safe to say that Giffen goods are very rare. The theory of consumer choice does allow demand curves to slope upward. Yet such occurrences are so unusual that the law of demand is as reliable a law as any in economics.

Giffen good

a good for which an increase in the price raises the quantity demanded

Figure 21-13

THE WORK-LEISURE DECISION. This figure shows Sally's budget constraint for deciding how much to work, her indifference curves for consumption and leisure, and her optimum.



HOW DO WAGES AFFECT LABOR SUPPLY?

So far we have used the theory of consumer choice to analyze how a person decides how to allocate his income between two goods. We can use the same theory to analyze how a person decides to allocate his time between work and leisure.

Consider the decision facing Sally, a freelance software designer. Sally is awake for 100 hours per week. She spends some of this time enjoying leisure—riding her bike, watching television, studying economics, and so on. She spends the rest of this time at her computer developing software. For every hour she spends developing software, she earns \$50, which she spends on consumption goods. Thus, her wage (\$50) reflects the tradeoff Sally faces between leisure and consumption. For every hour of leisure she gives up, she works one more hour and gets \$50 of consumption.

Figure 21-13 shows Sally's budget constraint. If she spends all 100 hours enjoying leisure, she has no consumption. If she spends all 100 hours working, she earns a weekly consumption of \$5,000 but has no time for leisure. If she works a normal 40-hour week, she enjoys 60 hours of leisure and has weekly consumption of \$2,000.

Figure 21-13 uses indifference curves to represent Sally's preferences for consumption and leisure. Here consumption and leisure are the two "goods" between which Sally is choosing. Because Sally always prefers more leisure and more consumption, she prefers points on higher indifference curves to points on lower ones. At a wage of \$50 per hour, Sally chooses a combination of consumption and leisure represented by the point labeled "optimum." This is the point on the budget constraint that is on the highest possible indifference curve, which is curve I_2 .

Now consider what happens when Sally's wage increases from \$50 to \$60 per hour. Figure 21-14 shows two possible outcomes. In each case, the budget constraint, shown in the left-hand graph, shifts outward from BC_1 to BC_2 . In the process, the budget constraint becomes steeper, reflecting the change in relative

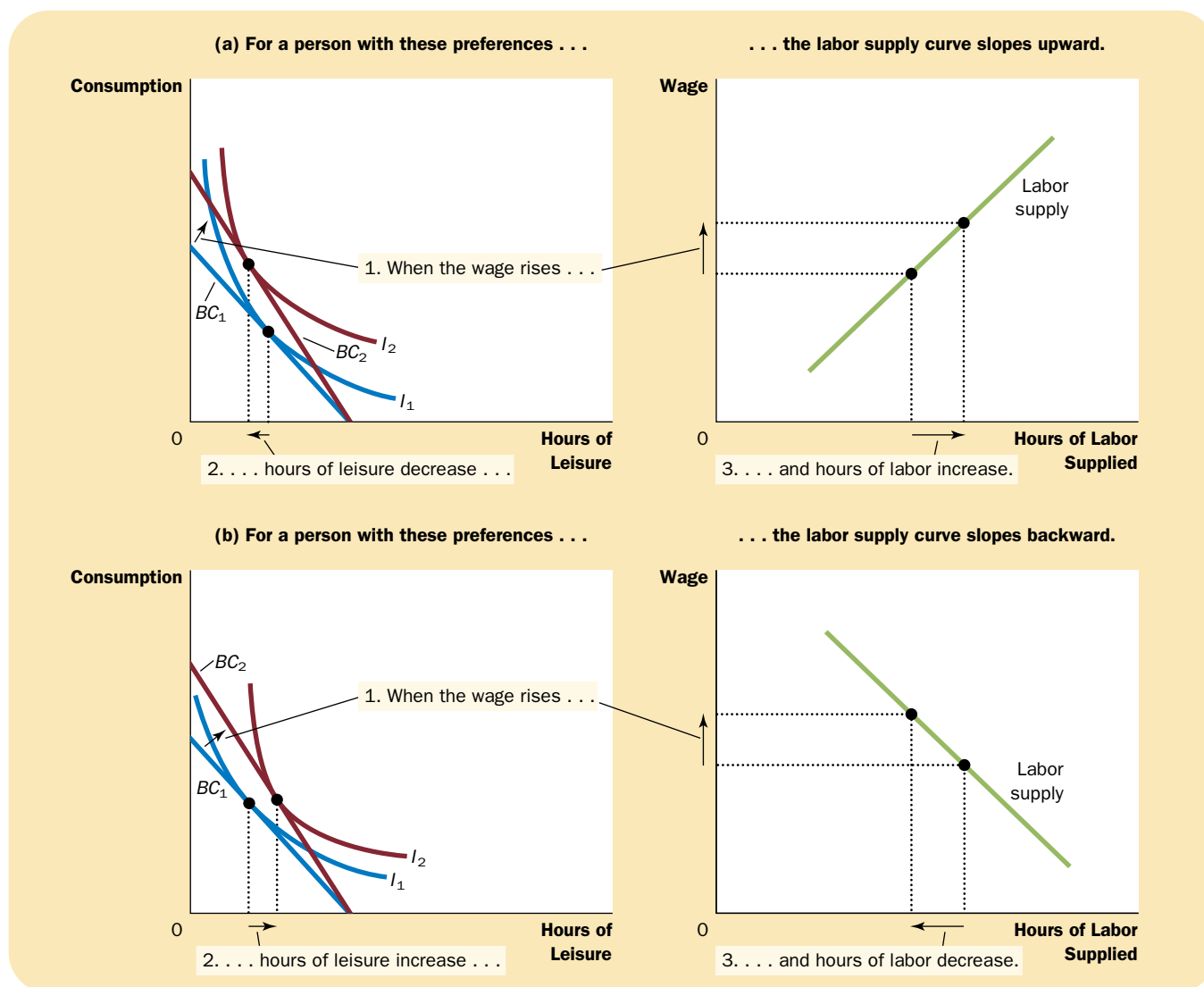


Figure 21-14

AN INCREASE IN THE WAGE. The two panels of this figure show how a person might respond to an increase in the wage. The graphs on the left show the consumer's initial budget constraint BC_1 and new budget constraint BC_2 , as well as the consumer's optimal choices over consumption and leisure. The graphs on the right show the resulting labor supply curve. Because hours worked equal total hours available minus hours of leisure, any change in leisure implies an opposite change in the quantity of labor supplied. In panel (a), when the wage rises, consumption rises and leisure falls, resulting in a labor supply curve that slopes upward. In panel (b), when the wage rises, both consumption and leisure rise, resulting in a labor supply curve that slopes backward.

price: At the higher wage, Sally gets more consumption for every hour of leisure that she gives up.

Sally's preferences, as represented by her indifference curves, determine the resulting responses of consumption and leisure to the higher wage. In both panels,

consumption rises. Yet the response of leisure to the change in the wage is different in the two cases. In panel (a), Sally responds to the higher wage by enjoying less leisure. In panel (b), Sally responds by enjoying more leisure.

Sally's decision between leisure and consumption determines her supply of labor, for the more leisure she enjoys the less time she has left to work. In each panel, the right-hand graph in Figure 21-14 shows the labor supply curve implied by Sally's decision. In panel (a), a higher wage induces Sally to enjoy less leisure and work more, so the labor supply curve slopes upward. In panel (b), a higher wage induces Sally to enjoy more leisure and work less, so the labor supply curve slopes "backward."

At first, the backward-sloping labor supply curve is puzzling. Why would a person respond to a higher wage by working less? The answer comes from considering the income and substitution effects of a higher wage.

Consider first the substitution effect. When Sally's wage rises, leisure becomes more costly relative to consumption, and this encourages Sally to substitute consumption for leisure. In other words, the substitution effect induces Sally to work harder in response to higher wages, which tends to make the labor supply curve slope upward.

Now consider the income effect. When Sally's wage rises, she moves to a higher indifference curve. She is now better off than she was. As long as consumption and leisure are both normal goods, she tends to want to use this increase in well-being to enjoy both higher consumption and greater leisure. In other words, the income effect induces her to work less, which tends to make the labor supply curve slope backward.

In the end, economic theory does not give a clear prediction about whether an increase in the wage induces Sally to work more or less. If the substitution effect is greater than the income effect for Sally, she works more. If the income effect is greater than the substitution effect, she works less. The labor supply curve, therefore, could be either upward or backward sloping.



"NO MORE 9-TO-5 FOR ME."

CASE STUDY INCOME EFFECTS ON LABOR SUPPLY: HISTORICAL TRENDS, LOTTERY WINNERS, AND THE CARNEGIE CONJECTURE

The idea of a backward-sloping labor supply curve might at first seem like a mere theoretical curiosity, but in fact it is not. Evidence indicates that the labor supply curve, considered over long periods of time, does in fact slope backward. A hundred years ago many people worked six days a week. Today five-day workweeks are the norm. At the same time that the length of the workweek has been falling, the wage of the typical worker (adjusted for inflation) has been rising.

Here is how economists explain this historical pattern: Over time, advances in technology raise workers' productivity and, thereby, the demand for labor. The increase in labor demand raises equilibrium wages. As wages rise, so does the reward for working. Yet rather than responding to this increased incentive by working more, most workers choose to take part of their greater prosperity in the form of more leisure. In other words, the income effect of higher wages dominates the substitution effect.

Further evidence that the income effect on labor supply is strong comes from a very different kind of data: winners of lotteries. Winners of large prizes

in the lottery see large increases in their incomes and, as a result, large outward shifts in their budget constraints. Because the winners' wages have not changed, however, the *slopes* of their budget constraints remain the same. There is, therefore, no substitution effect. By examining the behavior of lottery winners, we can isolate the income effect on labor supply.

The results from studies of lottery winners are striking. Of those winners who win more than \$50,000, almost 25 percent quit working within a year, and another 9 percent reduce the number of hours they work. Of those winners who win more than \$1 million, almost 40 percent stop working. The income effect on labor supply of winning such a large prize is substantial.

Similar results were found in a study, published in the May 1993 issue of the *Quarterly Journal of Economics*, of how receiving a bequest affects a person's labor supply. The study found that a single person who inherits more than \$150,000 is four times as likely to stop working as a single person who inherits less than \$25,000. This finding would not have surprised the nineteenth-century industrialist Andrew Carnegie. Carnegie warned that "the parent who leaves his son enormous wealth generally deadens the talents and energies of the son, and tempts him to lead a less useful and less worthy life than he otherwise would." That is, Carnegie viewed the income effect on labor supply to be substantial and, from his paternalistic perspective, regrettable. During his life and at his death, Carnegie gave much of his vast fortune to charity.

HOW DO INTEREST RATES AFFECT HOUSEHOLD SAVING?

An important decision that every person faces is how much income to consume today and how much to save for the future. We can use the theory of consumer choice to analyze how people make this decision and how the amount they save depends on the interest rate their savings will earn.

Consider the decision facing Sam, a worker planning ahead for retirement. To keep things simple, let's divide Sam's life into two periods. In the first period, Sam is young and working. In the second period, he is old and retired. When young, Sam earns a total of \$100,000. He divides this income between current consumption and saving. When he is old, Sam will consume what he has saved, including the interest that his savings have earned.

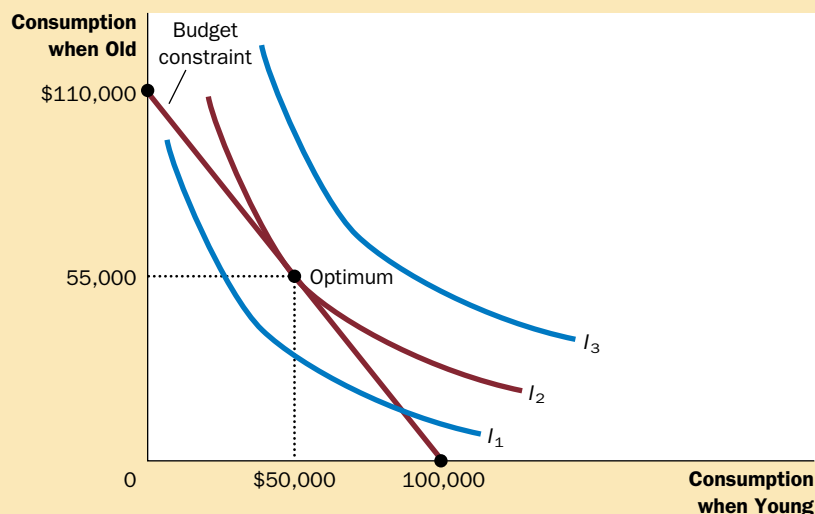
Suppose that the interest rate is 10 percent. Then for every dollar that Sam saves when young, he can consume \$1.10 when old. We can view "consumption when young" and "consumption when old" as the two goods that Sam must choose between. The interest rate determines the relative price of these two goods.

Figure 21-15 shows Sam's budget constraint. If he saves nothing, he consumes \$100,000 when young and nothing when old. If he saves everything, he consumes nothing when young and \$110,000 when old. The budget constraint shows these and all the intermediate possibilities.

Figure 21-15 uses indifference curves to represent Sam's preferences for consumption in the two periods. Because Sam prefers more consumption in both periods, he prefers points on higher indifference curves to points on lower ones. Given his preferences, Sam chooses the optimal combination of consumption in both periods of life, which is the point on the budget constraint that is on the highest possible indifference curve. At this optimum, Sam consumes \$50,000 when young and \$55,000 when old.

Figure 21-15

THE CONSUMPTION-SAVING DECISION. This figure shows the budget constraint for a person deciding how much to consume in the two periods of his life, the indifference curves representing his preferences, and the optimum.



Now consider what happens when the interest rate increases from 10 percent to 20 percent. Figure 21-16 shows two possible outcomes. In both cases, the budget constraint shifts outward and becomes steeper. At the new higher interest rate, Sam gets more consumption when old for every dollar of consumption that he gives up when young.

The two panels show different preferences for Sam and the resulting response to the higher interest rate. In both cases, consumption when old rises. Yet the response of consumption when young to the change in the interest rate is different in the two cases. In panel (a), Sam responds to the higher interest rate by consuming less when young. In panel (b), Sam responds by consuming more when young.

Sam's saving, of course, is his income when young minus the amount he consumes when young. In panel (a), consumption when young falls when the interest rate rises, so saving must rise. In panel (b), Sam consumes more when young, so saving must fall.

The case shown in panel (b) might at first seem odd: Sam responds to an increase in the return to saving by saving less. Yet this behavior is not as peculiar as it might seem. We can understand it by considering the income and substitution effects of a higher interest rate.

Consider first the substitution effect. When the interest rate rises, consumption when old becomes less costly relative to consumption when young. Therefore, the substitution effect induces Sam to consume more when old and less when young. In other words, the substitution effect induces Sam to save more.

Now consider the income effect. When the interest rate rises, Sam moves to a higher indifference curve. He is now better off than he was. As long as consumption in both periods consists of normal goods, he tends to want to use this increase in well-being to enjoy higher consumption in both periods. In other words, the income effect induces him to save less.

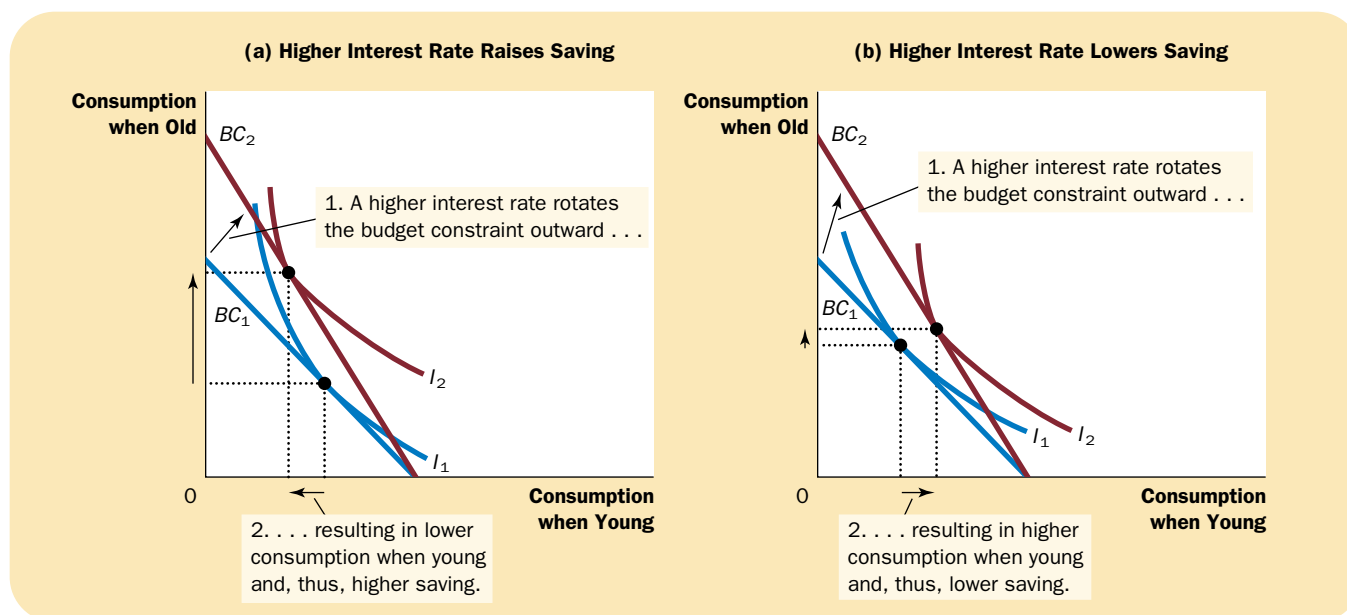


Figure 21-16

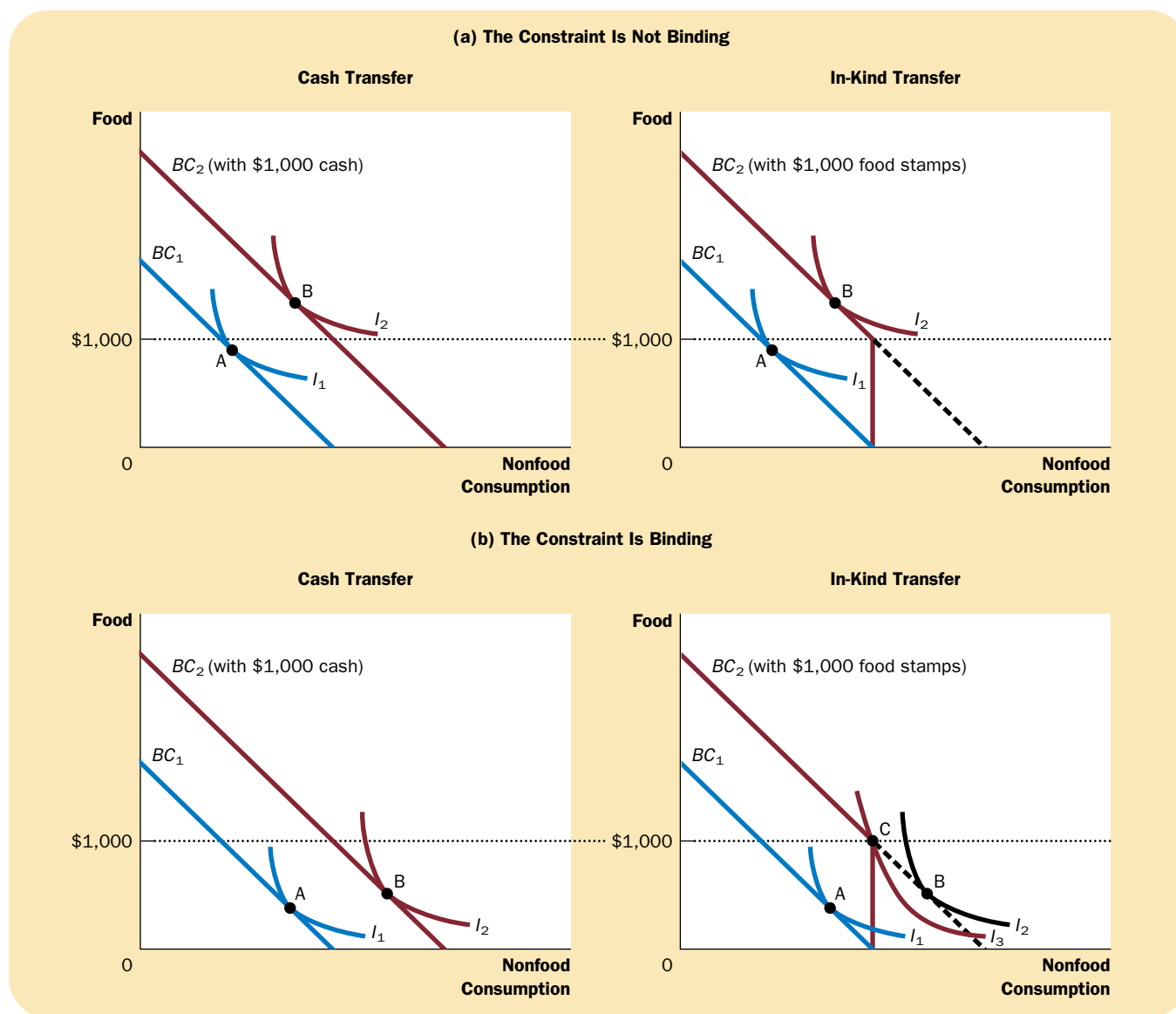
AN INCREASE IN THE INTEREST RATE. In both panels, an increase in the interest rate shifts the budget constraint outward. In panel (a), consumption when young falls, and consumption when old rises. The result is an increase in saving when young. In panel (b), consumption in both periods rises. The result is a decrease in saving when young.

The end result, of course, depends on both the income and substitution effects. If the substitution effect of a higher interest rate is greater than the income effect, Sam saves more. If the income effect is greater than the substitution effect, Sam saves less. Thus, the theory of consumer choice says that an increase in the interest rate could either encourage or discourage saving.

Although this ambiguous result is interesting from the standpoint of economic theory, it is disappointing from the standpoint of economic policy. It turns out that an important issue in tax policy hinges in part on how saving responds to interest rates. Some economists have advocated reducing the taxation of interest and other capital income, arguing that such a policy change would raise the after-tax interest rate that savers can earn and would thereby encourage people to save more. Other economists have argued that because of offsetting income and substitution effects, such a tax change might not increase saving and could even reduce it. Unfortunately, research has not led to a consensus about how interest rates affect saving. As a result, there remains disagreement among economists about whether changes in tax policy aimed to encourage saving would, in fact, have the intended effect.

DO THE POOR PREFER TO RECEIVE CASH OR IN-KIND TRANSFERS?

Paul is a pauper. Because of his low income, he has a meager standard of living. The government wants to help. It can either give Paul \$1,000 worth of food

**Figure 21-17**

CASH VERSUS IN-KIND TRANSFERS. Both panels compare a cash transfer and a similar in-kind transfer of food. In panel (a), the in-kind transfer does not impose a binding constraint, and the consumer ends up on the same indifference curve under the two policies. In panel (b), the in-kind transfer imposes a binding constraint, and the consumer ends up on a lower indifference curve with the in-kind transfer than with the cash transfer.

(perhaps by issuing him food stamps) or simply give him \$1,000 in cash. What does the theory of consumer choice have to say about the comparison between these two policy options?

Figure 21-17 shows how the two options might work. If the government gives Paul cash, then the budget constraint shifts outward. He can divide the extra cash

between food and nonfood consumption however he pleases. By contrast, if the government gives Paul an in-kind transfer of food, then his new budget constraint is more complicated. The budget constraint has again shifted out. But now the budget constraint has a kink at \$1,000 of food, for Paul must consume at least that amount in food. That is, even if Paul spends all his money on nonfood consumption, he still consumes \$1,000 in food.

The ultimate comparison between the cash transfer and in-kind transfer depends on Paul's preferences. In panel (a), Paul would choose to spend at least \$1,000 on food even if he receives a cash transfer. Therefore, the constraint imposed by the in-kind transfer is not binding. In this case, his consumption moves from point A to point B regardless of the type of transfer. That is, Paul's choice between food and nonfood consumption is the same under the two policies.

In panel (b), however, the story is very different. In this case, Paul would prefer to spend less than \$1,000 on food and spend more on nonfood consumption. The cash transfer allows him discretion to spend the money as he pleases, and he consumes at point B. By contrast, the in-kind transfer imposes the binding constraint that he consume at least \$1,000 of food. His optimal allocation is at the kink, point C. Compared to the cash transfer, the in-kind transfer induces Paul to consume more food and less of other goods. The in-kind transfer also forces Paul to end up on a lower (and thus less preferred) indifference curve. Paul is worse off than if he had the cash transfer.

Thus, the theory of consumer choice teaches a simple lesson about cash versus in-kind transfers. If an in-kind transfer of a good forces the recipient to consume more of the good than he would on his own, then the recipient prefers the cash transfer. If the in-kind transfer does not force the recipient to consume more of the good than he would on his own, then the cash and in-kind transfer have exactly the same effect on the consumption and welfare of the recipient.

QUICK QUIZ: Explain how an increase in the wage can potentially decrease the amount that a person wants to work.

CONCLUSION: DO PEOPLE REALLY THINK THIS WAY?

The theory of consumer choice describes how people make decisions. As we have seen, it has broad applicability. It can explain how a person chooses between Pepsi and pizza, work and leisure, consumption and saving, and on and on.

At this point, however, you might be tempted to treat the theory of consumer choice with some skepticism. After all, you are a consumer. You decide what to buy every time you walk into a store. And you know that you do not decide by writing down budget constraints and indifference curves. Doesn't this knowledge about your own decisionmaking provide evidence against the theory?

The answer is no. The theory of consumer choice does not try to present a literal account of how people make decisions. It is a model. And, as we first discussed in Chapter 2, models are not intended to be completely realistic.

The best way to view the theory of consumer choice is as a metaphor for how consumers make decisions. No consumer (except an occasional economist) goes

through the explicit optimization envisioned in the theory. Yet consumers are aware that their choices are constrained by their financial resources. And, given those constraints, they do the best they can to achieve the highest level of satisfaction. The theory of consumer choice tries to describe this implicit, psychological process in a way that permits explicit, economic analysis.

The proof of the pudding is in the eating. And the test of a theory is in its applications. In the last section of this chapter we applied the theory of consumer choice to four practical issues about the economy. If you take more advanced courses in economics, you will see that this theory provides the framework for much additional analysis.

Summary

- ◆ A consumer's budget constraint shows the possible combinations of different goods he can buy given his income and the prices of the goods. The slope of the budget constraint equals the relative price of the goods.
- ◆ The consumer's indifference curves represent his preferences. An indifference curve shows the various bundles of goods that make the consumer equally happy. Points on higher indifference curves are preferred to points on lower indifference curves. The slope of an indifference curve at any point is the consumer's marginal rate of substitution—the rate at which the consumer is willing to trade one good for the other.
- ◆ The consumer optimizes by choosing the point on his budget constraint that lies on the highest indifference curve. At this point, the slope of the indifference curve (the marginal rate of substitution between the goods) equals the slope of the budget constraint (the relative price of the goods).
- ◆ When the price of a good falls, the impact on the consumer's choices can be broken down into an income effect and a substitution effect. The income effect is the change in consumption that arises because a lower price makes the consumer better off. The substitution effect is the change in consumption that arises because a price change encourages greater consumption of the good that has become relatively cheaper. The income effect is reflected in the movement from a lower to a higher indifference curve, whereas the substitution effect is reflected by a movement along an indifference curve to a point with a different slope.
- ◆ The theory of consumer choice can be applied in many situations. It can explain why demand curves can potentially slope upward, why higher wages could either increase or decrease the quantity of labor supplied, why higher interest rates could either increase or decrease saving, and why the poor prefer cash to in-kind transfers.

Key Concepts

budget constraint, p. 465
 indifference curve, p. 466
 marginal rate of substitution, p. 467
 perfect substitutes, p. 470

perfect complements, p. 470
 normal good, p. 473
 inferior good, p. 473
 income effect, p. 475

substitution effect, p. 475
 Giffen good, p. 479

Questions for Review

1. A consumer has income of \$3,000. Wine costs \$3 a glass, and cheese costs \$6 a pound. Draw the consumer's budget constraint. What is the slope of this budget constraint?

2. Draw a consumer's indifference curves for wine and cheese. Describe and explain four properties of these indifference curves.
3. Pick a point on an indifference curve for wine and cheese and show the marginal rate of substitution. What does the marginal rate of substitution tell us?
4. Show a consumer's budget constraint and indifference curves for wine and cheese. Show the optimal consumption choice. If the price of wine is \$3 a glass and the price of cheese is \$6 a pound, what is the marginal rate of substitution at this optimum?
5. A person who consumes wine and cheese gets a raise, so his income increases from \$3,000 to \$4,000. Show what happens if both wine and cheese are normal goods. Now show what happens if cheese is an inferior good.
6. The price of cheese rises from \$6 to \$10 a pound, while the price of wine remains \$3 a glass. For a consumer with a constant income of \$3,000, show what happens to consumption of wine and cheese. Decompose the change into income and substitution effects.
7. Can an increase in the price of cheese possibly induce a consumer to buy more cheese? Explain.
8. Suppose a person who buys only wine and cheese is given \$1,000 in food stamps to supplement his \$1,000 income. The food stamps cannot be used to buy wine. Might the consumer be better off with \$2,000 in income? Explain in words and with a diagram.

Problems and Applications

1. Jennifer divides her income between coffee and croissants (both of which are normal goods). An early frost in Brazil causes a large increase in the price of coffee in the United States.
 - a. Show the effect of the frost on Jennifer's budget constraint.
 - b. Show the effect of the frost on Jennifer's optimal consumption bundle assuming that the substitution effect outweighs the income effect for croissants.
 - c. Show the effect of the frost on Jennifer's optimal consumption bundle assuming that the income effect outweighs the substitution effect for croissants.
2. Compare the following two pairs of goods:
 - ◆ Coke and Pepsi
 - ◆ Skis and ski bindings

In which case do you expect the indifference curves to be fairly straight, and in which case do you expect the indifference curves to be very bowed? In which case will the consumer respond more to a change in the relative price of the two goods?
3. Mario consumes only cheese and crackers.
 - a. Could cheese and crackers both be inferior goods for Mario? Explain.
 - b. Suppose that cheese is a normal good for Mario whereas crackers are an inferior good. If the price of cheese falls, what happens to Mario's consumption of crackers? What happens to his consumption of cheese? Explain.
4. Jim buys only milk and cookies.
 - a. In 2001, Jim earns \$100, milk costs \$2 per quart, and cookies cost \$4 per dozen. Draw Jim's budget constraint.
 - b. Now suppose that all prices increase by 10 percent in 2002 and that Jim's salary increases by 10 percent as well. Draw Jim's new budget constraint. How would Jim's optimal combination of milk and cookies in 2002 compare to his optimal combination in 2001?
5. Consider your decision about how many hours to work.
 - a. Draw your budget constraint assuming that you pay no taxes on your income. On the same diagram, draw another budget constraint assuming that you pay a 15 percent tax.
 - b. Show how the tax might lead to more hours of work, fewer hours, or the same number of hours. Explain.
6. Sarah is awake for 100 hours per week. Using one diagram, show Sarah's budget constraints if she earns \$6 per hour, \$8 per hour, and \$10 per hour. Now draw indifference curves such that Sarah's labor supply curve is upward sloping when the wage is between \$6 and \$8 per hour, and backward sloping when the wage is between \$8 and \$10 per hour.
7. Draw the indifference curve for someone deciding how much to work. Suppose the wage increases. Is it possible that the person's consumption would fall? Is this

plausible? Discuss. (Hint: Think about income and substitution effects.)

8. Suppose you take a job that pays \$30,000 and set some of this income aside in a savings account that pays an annual interest rate of 5 percent. Use a diagram with a budget constraint and indifference curves to show how your consumption changes in each of the following situations. To keep things simple, assume that you pay no taxes on your income.
 - a. Your salary increases to \$40,000.
 - b. The interest rate on your bank account rises to 8 percent.
9. As discussed in the text, we can divide an individual's life into two hypothetical periods: "young" and "old." Suppose that the individual earns income only when young and saves some of that income to consume when old. If the interest rate on savings falls, can you tell what happens to consumption when young? Can you tell what happens to consumption when old? Explain.
10. Suppose that your state gives each town \$5 million in aid per year. The way in which the money is spent is currently unrestricted, but the governor has proposed that towns be required to spend the entire \$5 million on education. You can illustrate the effect of this proposal on your town's spending on education using a budget constraint and indifference-curve diagram. The two goods are education and noneducation spending.
 - a. Draw your town's budget constraint under the existing policy, assuming that your town's only source of revenue besides the state aid is a property tax that yields \$10 million. On the same diagram, draw the budget constraint under the governor's proposal.
 - b. Would your town spend more on education under the governor's proposal than under the existing policy? Explain.
 - c. Now compare two towns—Youngsville and Oldsville—with the same revenue and the same state aid. Youngsville has a large school-age population, and Oldsville has a large elderly population. In which town is the governor's proposal most likely to increase education spending? Explain.
11. (This problem is challenging.) The welfare system provides income to some needy families. Typically, the maximum payment goes to families that earn no income; then, as families begin to earn income, the welfare payment declines gradually and eventually disappears. Let's consider the possible effects of this program on a family's labor supply.
 - a. Draw a budget constraint for a family assuming that the welfare system did not exist. On the same diagram, draw a budget constraint that reflects the existence of the welfare system.
 - b. Adding indifference curves to your diagram, show how the welfare system could reduce the number of hours worked by the family. Explain, with reference to both the income and substitution effects.
 - c. Using your diagram from part (b), show the effect of the welfare system on the well-being of the family.
12. (This problem is challenging.) Suppose that an individual owed no taxes on the first \$10,000 she earned and 15 percent of any income she earned over \$10,000. (This is a simplified version of the actual U.S. income tax.) Now suppose that Congress is considering two ways to reduce the tax burden: a reduction in the tax rate and an increase in the amount on which no tax is owed.
 - a. What effect would a reduction in the tax rate have on the individual's labor supply if she earned \$30,000 to start? Explain in words using the income and substitution effects. You do not need to use a diagram.
 - b. What effect would an increase in the amount on which no tax is owed have on the individual's labor supply? Again, explain in words using the income and substitution effects.
13. (This problem is challenging.) Consider a person deciding how much to consume and how much to save for retirement. This person has particular preferences: Her lifetime utility depends on the lowest level of consumption during the two periods of her life. That is,

$$\text{Utility} = \text{Minimum} \{ \text{consumption when young, consumption when old} \}.$$
 - a. Draw this person's indifference curves. (Hint: Recall that indifference curves show the combinations of consumption in the two periods that yield the same level of utility.)
 - b. Draw the budget constraint and the optimum.
 - c. When the interest rate increases, does this person save more or less? Explain your answer using income and substitution effects.

GLOSSARY

ability-to-pay principle—the idea that taxes should be levied on a person according to how well that person can shoulder the burden

absolute advantage—the comparison among producers of a good according to their productivity

accounting profit—total revenue minus total explicit cost

average fixed cost—fixed costs divided by the quantity of output

average revenue—total revenue divided by the quantity sold

average tax rate—total taxes paid divided by total income

average total cost—total cost divided by the quantity of output

average variable cost—variable costs divided by the quantity of output

benefits principle—the idea that people should pay taxes based on the benefits they receive from government services

budget constraint—the limit on the consumption bundles that a consumer can afford

budget deficit—a shortfall of tax revenue from government spending

budget surplus—an excess of government receipts over government spending

capital—the equipment and structures used to produce goods and services

cartel—a group of firms acting in unison

ceteris paribus—a Latin phrase, translated as “other things being equal,” used as a reminder that all variables other than the ones being studied are assumed to be constant

circular-flow diagram—a visual model of the economy that shows how dollars flow through markets among households and firms

Coase theorem—the proposition that if private parties can bargain without cost over the allocation of resources, they can solve the problem of externalities on their own

collusion—an agreement among firms in a market about quantities to produce or prices to charge

common resources—goods that are rival but not excludable

comparable worth—a doctrine according to which jobs deemed comparable should be paid the same wage

comparative advantage—the comparison among producers of a good according to their opportunity cost

compensating differential—a difference in wages that arises to offset the nonmonetary characteristics of different jobs

competitive market—a market with many buyers and sellers trading identical products so that each buyer and seller is a price taker

complements—two goods for which an increase in the price of one leads to a decrease in the demand for the other

constant returns to scale—the property whereby long-run average total cost stays the same as the quantity of output changes

consumer surplus—a buyer’s willingness to pay minus the amount the buyer actually pays

cost—the value of everything a seller must give up to produce a good

cost-benefit analysis—a study that compares the costs and benefits to society of providing a public good

cross-price elasticity of demand—a measure of how much the quantity demanded of one good responds to a change in the price of another good, computed as the percentage change in quantity demanded of the first good divided by the percentage change in the price

deadweight loss—the fall in total surplus that results from a market distortion, such as a tax

demand curve—a graph of the relationship between the price of a good and the quantity demanded

demand schedule—a table that shows the relationship between the price of a good and the quantity demanded

diminishing marginal product—the property whereby the marginal product of an input declines as the quantity of the input increases

discrimination—the offering of different opportunities to similar individuals who differ only by race, ethnic group, sex, age, or other personal characteristics

diseconomies of scale—the property whereby long-run average total cost rises as the quantity of output increases

dominant strategy—a strategy that is best for a player in a game regardless of the strategies chosen by the other players

economic profit—total revenue minus total cost, including both explicit and implicit costs

economics—the study of how society manages its scarce resources

economies of scale—the property whereby long-run average total cost falls as the quantity of output increases

efficiency—the property of society getting the most it can from its scarce resources

efficiency wages—above-equilibrium wages paid by firms in order to increase worker productivity

efficient scale—the quantity of output that minimizes average total cost

elasticity—a measure of the responsiveness of quantity demanded or quantity supplied to one of its determinants

equilibrium—a situation in which supply and demand have been brought into balance

equilibrium price—the price that balances supply and demand

equilibrium quantity—the quantity supplied and the quantity demanded when the price has adjusted to balance supply and demand

equity—the property of distributing economic prosperity fairly among the members of society

excludability—the property of a good whereby a person can be prevented from using it

explicit costs—input costs that require an outlay of money by the firm

exports—goods and services that are produced domestically and sold abroad

externality—the impact of one person’s actions on the well-being of a bystander

factors of production—the inputs used to produce goods and services

fixed costs—costs that do not vary with the quantity of output produced

free rider—a person who receives the benefit of a good but avoids paying for it

game theory—the study of how people behave in strategic situations

Giffen good—a good for which an increase in the price raises the quantity demanded

horizontal equity—the idea that taxpayers with similar abilities to pay taxes should pay the same amount

human capital—the accumulation of investments in people, such as education and on-the-job training

implicit costs—input costs that do not require an outlay of money by the firm

import quota—a limit on the quantity of a good that can be produced abroad and sold domestically

imports—goods and services that are produced abroad and sold domestically

in-kind transfers—transfers to the poor given in the form of goods and services rather than cash

income effect—the change in consumption that results when a price change moves the consumer to a higher or lower indifference curve

income elasticity of demand—a measure of how much the quantity demanded of a good responds to a change in consumers' income, computed as the percentage change in quantity demanded divided by the percentage change in income

indifference curve—a curve that shows consumption bundles that give the consumer the same level of satisfaction

inferior good—a good for which, other things equal, an increase in income leads to a decrease in demand

inflation—an increase in the overall level of prices in the economy

internalizing an externality—altering incentives so that people take account of the external effects of their actions

law of demand—the claim that, other things equal, the quantity demanded of a good falls when the price of the good rises

law of supply—the claim that, other things equal, the quantity supplied of a good rises when the price of the good rises

law of supply and demand—the claim that the price of any good adjusts to

bring the supply and demand for that good into balance

liberalism—the political philosophy according to which the government should choose policies deemed to be just, as evaluated by an impartial observer behind a "veil of ignorance"

libertarianism—the political philosophy according to which the government should punish crimes and enforce voluntary agreements but not redistribute income

life cycle—the regular pattern of income variation over a person's life

lump-sum tax—a tax that is the same amount for every person

macroeconomics—the study of economy-wide phenomena, including inflation, unemployment, and economic growth

marginal changes—small incremental adjustments to a plan of action

marginal cost—the increase in total cost that arises from an extra unit of production

marginal product—the increase in output that arises from an additional unit of input

marginal product of labor—the increase in the amount of output from an additional unit of labor

marginal rate of substitution—the rate at which a consumer is willing to trade one good for another

marginal revenue—the change in total revenue from an additional unit sold

marginal tax rate—the extra taxes paid on an additional dollar of income

market—a group of buyers and sellers of a particular good or service

market economy—an economy that allocates resources through the decentralized decisions of many firms and households as they interact in markets for goods and services

market failure—a situation in which a market left on its own fails to allocate resources efficiently

market power—the ability of a single economic actor (or small group of actors) to have a substantial influence on market prices

maximin criterion—the claim that the government should aim to

maximize the well-being of the worst-off person in society

microeconomics—the study of how households and firms make decisions and how they interact in markets

monopolistic competition—a market structure in which many firms sell products that are similar but not identical

monopoly—a firm that is the sole seller of a product without close substitutes

Nash equilibrium—a situation in which economic actors interacting with one another each choose their best strategy given the strategies that all the other actors have chosen

natural monopoly—a monopoly that arises because a single firm can supply a good or service to an entire market at a smaller cost than could two or more firms

negative income tax—a tax system that collects revenue from high-income households and gives transfers to low-income households

normal good—a good for which, other things equal, an increase in income leads to an increase in demand

normative statements—claims that attempt to prescribe how the world should be

oligopoly—a market structure in which only a few sellers offer similar or identical products

opportunity cost—whatever must be given up to obtain some item

perfect complements—two goods with right-angle indifference curves

perfect substitutes—two goods with straight-line indifference curves

permanent income—a person's normal income

Phillips curve—a curve that shows the short-run tradeoff between inflation and unemployment

Pigovian tax—a tax enacted to correct the effects of a negative externality

positive statements—claims that attempt to describe the world as it is

poverty line—an absolute level of income set by the federal government for each family size below which a family is deemed to be in poverty

poverty rate—the percentage of the population whose family income

- falls below an absolute level called the poverty line
- price ceiling**—a legal maximum on the price at which a good can be sold
- price discrimination**—the business practice of selling the same good at different prices to different customers
- price elasticity of demand**—a measure of how much the quantity demanded of a good responds to a change in the price of that good, computed as the percentage change in quantity demanded divided by the percentage change in price
- price elasticity of supply**—a measure of how much the quantity supplied of a good responds to a change in the price of that good, computed as the percentage change in quantity supplied divided by the percentage change in price
- price floor**—a legal minimum on the price at which a good can be sold
- prisoners' dilemma**—a particular "game" between two captured prisoners that illustrates why cooperation is difficult to maintain even when it is mutually beneficial
- private goods**—goods that are both excludable and rival
- producer surplus**—the amount a seller is paid for a good minus the seller's cost
- production function**—the relationship between quantity of inputs used to make a good and the quantity of output of that good
- production possibilities frontier**—a graph that shows the combinations of output that the economy can possibly produce given the available factors of production and the available production technology
- productivity**—the amount of goods and services produced from each hour of a worker's time
- profit**—total revenue minus total cost
- progressive tax**—a tax for which high-income taxpayers pay a larger fraction of their income than do low-income taxpayers
- proportional tax**—a tax for which high-income and low-income taxpayers pay the same fraction of income
- public goods**—goods that are neither excludable nor rival
- quantity demanded**—the amount of a good that buyers are willing and able to purchase
- quantity supplied**—the amount of a good that sellers are willing and able to sell
- regressive tax**—a tax for which high-income taxpayers pay a smaller fraction of their income than do low-income taxpayers
- rivalry**—the property of a good whereby one person's use diminishes other people's use
- scarcity**—the limited nature of society's resources
- shortage**—a situation in which quantity demanded is greater than quantity supplied
- strike**—the organized withdrawal of labor from a firm by a union
- substitutes**—two goods for which an increase in the price of one leads to an increase in the demand for the other
- substitution effect**—the change in consumption that results when a price change moves the consumer along a given indifference curve to a point with a new marginal rate of substitution
- sunk cost**—a cost that has already been committed and cannot be recovered
- supply curve**—a graph of the relationship between the price of a good and the quantity supplied
- supply schedule**—a table that shows the relationship between the price of a good and the quantity supplied
- surplus**—a situation in which quantity supplied is greater than quantity demanded
- tariff**—a tax on goods produced abroad and sold domestically
- tax incidence**—the study of who bears the burden of taxation
- total cost**—the market value of the inputs a firm uses in production
- total revenue (for a firm)**—the amount a firm receives for the sale of its output
- total revenue (in a market)**—the amount paid by buyers and received by sellers of a good, computed as the price of the good times the quantity sold
- Tragedy of the Commons**—a parable that illustrates why common resources get used more than is desirable from the standpoint of society as a whole
- transaction costs**—the costs that parties incur in the process of agreeing and following through on a bargain
- union**—a worker association that bargains with employers over wages and working conditions
- utilitarianism**—the political philosophy according to which the government should choose policies to maximize the total utility of everyone in society
- utility**—a measure of happiness or satisfaction
- value of the marginal product**—the marginal product of an input times the price of the output
- variable costs**—costs that do vary with the quantity of output produced
- vertical equity**—the idea that taxpayers with a greater ability to pay taxes should pay larger amounts
- welfare**—government programs that supplement the incomes of the needy
- welfare economics**—the study of how the allocation of resources affects economic well-being
- willingness to pay**—the maximum amount that a buyer will pay for a good
- world price**—the price of a good that prevails in the world market for that good

