Principles of Microeconomics

Course Reader

August 1, 2013

This is a course reader assembled by Professor Matthew J. Holian. It contains chapters from Rittenberg and Tregarthen's *Principles of Economics*.

References:

Rittenberg, Libby & Timothy Tregarthen, 2009. *Principles of Economics*. 1st edition. Flat World Knowledge (www.flatworldknowledge.com).

Table of Contents (*Brief*)

Table of Contents (*Detailed*)

CHAPTER 1

Economics: The Study of Choice

START UP: ECONOMICS IN THE NEWS

2008 seemed to be the year of economic news. From the worst financial crisis since the Great Depression to the possibility of a global recession, to gyrating gasoline and food prices, and to plunging housing prices, economic questions were the primary factors in the presidential campaign of 2008 and dominated the news generally.

What causes the prices of some good to rise while the prices of some other goods fall? Price determination is one of the things that we will study in this book. We will also consider factors that lead an economy to fall into a recession—and the attempts to limit it.

While the investigation of these problems surely falls within the province of economics, economics encompasses a far broader range of issues. Ultimately, economics is the study of choice. Because choices range over every imaginable aspect of human experience, so does economics. Economists have investigated the nature of family life, the arts, education, crime, sports, job creation—the list is virtually endless because so much of our lives involves making choices.

How do individuals make choices: Would you like better grades? More time to relax? More time watching movies? Getting better grades probably requires more time studying, and perhaps less relaxation and entertainment. Not only must we make choices as individuals, we must make choices as a society. Do we want a cleaner environment? Faster economic growth? Both may be desirable, but efforts to clean up the environment may conflict with faster economic growth. Society must make choices.

Economics is defined less by the subjects economists investigate than by the way in which economists investigate them. Economists have a way of looking at the world that differs from the way scholars in other disciplines look at the world. It is the *economic way of thinking*; this chapter introduces that way of thinking.

1. DEFINING ECONOMICS

LEARNING OBJECTIVES

1. Define economics.
2. Explain the concepts of scarcity and opportunity cost and how they relate to the definition of economics.
3. Understand the three fundamental economic questions: What should be produced? How should goods and services be produced? For whom should goods and services be produced?

economics

A social science that examines how people choose among the alternatives available to them.

Economics is a social science that examines how people choose among the alternatives available to them. It is social because it involves people and their behavior. It is a science because it uses, as much as possible, a scientific approach in its investigation of choices.

1.1 Scarcity, Choice, and Cost

All choices mean that one alternative is selected over another. Selecting among alternatives involves three ideas central to economics: scarcity, choice, and opportunity cost.

Scarcity

Our resources are limited. At any one time, we have only so much land, so many factories, so much oil, so many people. But our wants, our desires for the things that we can produce with those resources, are unlimited. We would always like more and better housing, more and better education—more and better of practically everything.

If our resources were also unlimited, we could say yes to each of our wants—and there would be no economics. Because our resources are limited, we cannot say yes to everything. To say yes to one thing requires that we say no to another. Whether we like it or not, we must make choices.

scarcity

The condition of having to choose among alternatives.

scarce good

A good for which the choice of one alternative requires that another be given up.

Our unlimited wants are continually colliding with the limits of our resources, forcing us to pick some activities and to reject others. **Scarcity** is the condition of having to choose among alternatives. A **scarce good** is one for which the choice of one alternative requires that another be given up.

Consider a parcel of land. The parcel presents us with several alternative uses. We could build a house on it. We could put a gas station on it. We could create a small park on it. We could leave the land undeveloped in order to be able to make a decision later as to how it should be used.

Suppose we have decided the land should be used for housing. Should it be a large and expensive house or several modest ones? Suppose it is to be a large and expensive house. Who should live in the house? If the Lees live in it, the Nguyens cannot. There are alternative uses of the land both in the sense of the type of use and also in the sense of who gets to use it. The fact that land is scarce means that society must make choices concerning its use.

Virtually everything is scarce. Consider the air we breathe, which is available in huge quantity at no charge to us. Could it possibly be scarce?

The test of whether air is scarce is whether it has alternative uses. What uses can we make of the air? We breathe it. We pollute it when we drive our cars, heat our houses, or operate our factories. In effect, one use of the air is as a garbage dump. We certainly need the air to breathe. But just as certainly, we choose to dump garbage in it. Those two uses are clearly alternatives to each other. The more garbage we dump in the air, the less desirable—and healthy—it will be to breathe. If we decide we want to breathe cleaner air, we must limit the activities that generate pollution. Air is a scarce good because it has alternative uses.

free good

A good for which the choice of one use does not require that another be given up.

Not all goods, however, confront us with such choices. A **free good** is one for which the choice of one use does not require that we give up another. One example of a free good is gravity. The fact that gravity is holding you to the earth does not mean that your neighbor is forced to drift up into space! One person's use of gravity is not an alternative to another person's use.

There are not many free goods. Outer space, for example, was a free good when the only use we made of it was to gaze at it. But now, our use of space has reached the point where one use can be an alternative to another. Conflicts have already arisen over the allocation of orbital slots for communications satellites. Thus, even parts of outer space are scarce. Space will surely become more scarce as we find new ways to use it. Scarcity characterizes virtually everything. Consequently, the scope of economics is wide indeed.

Scarcity and the Fundamental Economic Questions

The choices we confront as a result of scarcity raise three sets of issues. Every economy must answer the following questions:

1. **What should be produced?** Using the economy's scarce resources to produce one thing requires giving up another. Producing better education, for example, may require cutting back on other services, such as health care. A decision to preserve a wilderness area requires giving up other uses of the land. Every society must decide what it will produce with its scarce resources.
2. **How should goods and services be produced?** There are all sorts of choices to be made in determining how goods and services should be produced. Should a firm employ a few skilled or a lot of unskilled workers? Should it produce in its own country or should it use foreign plants? Should manufacturing firms use new or recycled raw materials to make their products?
3. **For whom should goods and services be produced?** If a good or service is produced, a decision must be made about who will get it. A decision to have one person or group receive a good or service usually means it will not be available to someone else. For example, representatives of the poorest nations on earth often complain that energy consumption per person in the United States is 17 *times* greater than energy consumption per person in the world's 62 poorest countries. Critics argue that the world's energy should be more evenly allocated. Should it? That is a "for whom" question.

Every economy must determine what should be produced, how it should be produced, and for whom it should be produced. We shall return to these questions again and again.

Opportunity Cost

It is within the context of scarcity that economists define what is perhaps the most important concept in all of economics, the concept of opportunity cost. **Opportunity cost** is the value of the best alternative forgone in making any choice.

opportunity cost

The value of the best alternative forgone in making any choice.

The opportunity cost to you of reading the remainder of this chapter will be the value of the best other use to which you could have put your time. If you choose to spend $20 on a potted plant, you have simultaneously chosen to give up the benefits of spending the $20 on pizzas or a paperback book or a night at the movies. If the book is the most valuable of those alternatives, then the opportunity cost of the plant is the value of the enjoyment you otherwise expected to receive from the book.

The concept of opportunity cost must not be confused with the purchase price of an item. Consider the cost of a college or university education. That includes the value of the best alternative use of money spent for tuition, fees, and books. But the most important cost of a college education is the value of the forgone alternative uses of time spent studying and attending class instead of using the time in some other endeavor. Students sacrifice that time in hopes of even greater earnings in the future or because they place a value on the opportunity to learn. Or consider the cost of going to the doctor. Part of that cost is the value of the best alternative use of the money required to see the doctor. But, the cost also includes the value of the best alternative use of the time required to see the doctor. The essential thing to see in the concept of opportunity cost is found in the name of the concept. Opportunity cost is the value of the best opportunity forgone in a particular choice. It is not simply the amount spent on that choice.

The concepts of scarcity, choice, and opportunity cost are at the heart of economics. A good is scarce if the choice of one alternative requires that another be given up. The existence of alternative uses forces us to make choices. The opportunity cost of any choice is the value of the best alternative forgone in making it.

KEY TAKEAWAYS

- Economics is a social science that examines how people choose among the alternatives available to them.
- Scarcity implies that we must give up one alternative in selecting another. A good that is not scarce is a free good.
- The three fundamental economic questions are: What should be produced? How should goods and services be produced? For whom should goods and services be produced?
- Every choice has an opportunity cost and opportunity costs affect the choices people make. The opportunity cost of any choice is the value of the best alternative that had to be forgone in making that choice.

TRY IT!

Identify the elements of scarcity, choice, and opportunity cost in each of the following:

1. The Environmental Protection Agency is considering an order that a 500-acre area on the outskirts of a large city be preserved in its natural state, because the area is home to a rodent that is considered an endangered species. Developers had planned to build a housing development on the land.
2. The manager of an automobile assembly plant is considering whether to produce cars or sport utility vehicles (SUVs) next month. Assume that the quantities of labor and other materials required would be the same for either type of production.
3. A young man who went to work as a nurses' aide after graduating from high school leaves his job to go to college, where he will obtain training as a registered nurse.

Case in Point: The Rising Cost of Energy

Oil is an exhaustible resource. The oil we burn today will not be available for use in the future. Part of the opportunity cost of our consumption of goods such as gasoline that are produced from oil includes the value people in the future might have placed on oil we use today.

It appears that the cost of our use of oil may be rising. We have been using "light crude," the oil found in the ground in deposits that can be readily tapped. As light crude becomes more scarce, the world may need to turn to so-called "heavy crude," the crude oil that is found in the sandy soil of places such as Canada and Venezuela. That oil exists in such abundance that it propels Venezuela to the top of the world list of available oil. Saudi Arabia moves to the second position; Canada is third.

The difficulty with the oil mixed in the sand is that extracting it is far more costly than light crude, both in terms of the expenditures required and in terms of the environmental damage that mining it creates. Northern Alberta, in Canada, boasts a Florida-sized area whose sandy soils are rich in crude oil. Some of that oil is 1,200 feet underground. Extracting it requires pumping steam into the oily sand and then pumping up the resultant oily syrup. That syrup is then placed into huge, industrial-sized washing machines that separate crude oil. What is left over is toxic and will be placed in huge lakes that are being created by digging pits in the ground 200 feet deep. The oil produced from these sands has become important—Alberta is the largest foreign supplier of oil to the United States.

Sands that are closer to the surface are removed by bulldozers and giant cranes; the forest over it is cleared away. The oily sand is then hauled off in two-story dump trucks which, when filled, weigh more than a Boeing 747. Total SA, a French company, is leading the race to develop Canada's oil. Jean Luc-Guiziou, the president of Total SA's Canadian operations, says that the extraordinarily costly process of extracting heavy crude is something the world is going to have to get used to. "The light crude undiscovered today is getting scarcer and scarcer," he told *The Wall Street Journal*. "We have to accept the reality of geoscience, which is that the next generation of oil resources will be heavier."

Already, Total SA has clear-cut thousands of acres of forest land in order to gain access to the oily sand below. The process of extracting heavy crude oil costs the company $25 a barrel—compared to the $6 per barrel cost of extracting and refining light crude. Extracting heavy crude generates three times as much greenhouse gas per barrel as does light crude. By 2015, Fort McMurray, the small (population 61,000) town that has become the headquarters of Northern Alberta's crude oil boom, will emit more greenhouse gas than the entire country of Denmark (population 5.4 million). Canada will exceed its greenhouse gas quota set by the Kyoto Accords—an international treaty aimed at limiting global warming—largely as a result of developing its heavy crude deposits.

No one even considered the extraction of heavy crude when light crude was cheap. In the late 1990s, oil cost just $12 per barrel, and deposits of heavy crude such as those in Canada attracted little attention. By mid-2006, oil sold for more than $70 per barrel, and Canada's heavy crude was suddenly a hot commodity. "It moved from being just an interesting experiment in northern Canada to really this is the future source of oil supply," Greg Stringham of the Canadian Association of Petroleum Producers told Al Jazeera.

Alberta's energy minister, Greg Melchin, defends the province's decision to proceed with the exploitation of its oily sand. "There is a cost to it, but the benefits are substantially greater," he insists.

Not everyone agrees. George Poitras, a member of the Mikisew Cree tribe, lives downstream from the oil sands development. "You see a lot of the land dug up, a lot of the boreal forest struck down and it's upsetting, it fills me with rage," he says. Diana Gibson of the Parkland Institute, an environmental advocacy group, says that you can see the environmental damage generated by the extraction of oil sands around Fort McMurray from the moon. "What we are going to be having is destruction of very, very valuable ecosystems, and permanent pollution," she says.

Sources: "Alberta's Heavy Oil Burden," Al Jazeera English, March 17, 2008 (see english.aljazeera.net); and Russell Gold, "As Prices Surge, Oil Giants Turn Sludge into Gold," The Wall Street Journal Online, March 27, 2006, A1.

ANSWERS TO TRY IT! PROBLEMS

1. The 500-acre area is scarce because it has alternative uses: preservation in its natural state or a site for homes. A choice must be made between these uses. The opportunity cost of preserving the land in its natural state is the forgone value of the land as a housing development. The opportunity cost of using the land as a housing development is the forgone value of preserving the land.
2. The scarce resources are the plant and the labor at the plant. The manager must choose between producing cars and producing SUVs. The opportunity cost of producing cars is the profit that could be earned from producing SUVs; the opportunity cost of producing SUVs is the profit that could be earned from producing cars.
3. The man can devote his time to his current career or to an education; his time is a scarce resource. He must choose between these alternatives. The opportunity cost of continuing as a nurses' aide is the forgone benefit he expects from training as a registered nurse; the opportunity cost of going to college is the forgone income he could have earned working full-time as a nurses' aide.

2. THE FIELD OF ECONOMICS

LEARNING OBJECTIVES

1. Explain the distinguishing characteristics of the economic way of thinking.
2. Distinguish between microeconomics and macroeconomics.

We have examined the basic concepts of scarcity, choice, and opportunity cost in economics. In this section, we will look at economics as a field of study. We begin with the characteristics that distinguish economics from other social sciences.

2.1 The Economic Way of Thinking

Economists study choices that scarcity requires us to make. This fact is not what distinguishes economics from other social sciences; all social scientists are interested in choices. An anthropologist might study the choices of ancient peoples; a political scientist might study the choices of legislatures; a psychologist might study how people choose a mate; a sociologist might study the factors that have led to a rise in single-parent households. Economists study such questions as well. What is it about the study of choices by economists that makes economics different from these other social sciences?

Three features distinguish the economic approach to choice from the approaches taken in other social sciences:

1. Economists give special emphasis to the role of opportunity costs in their analysis of choices.
2. Economists assume that individuals make choices that seek to maximize the value of some objective, and that they define their objectives in terms of their own self-interest.
3. Individuals maximize by deciding whether to do a little more or a little less of something. Economists argue that individuals pay attention to the consequences of small changes in the levels of the activities they pursue.

The emphasis economists place on opportunity cost, the idea that people make choices that maximize the value of objectives that serve their self-interest, and a focus on the effects of small changes are ideas of great power. They constitute the core of economic thinking. The next three sections examine these ideas in greater detail.

Opportunity Costs Are Important

If doing one thing requires giving up another, then the expected benefits of the alternatives we face will affect the ones we choose. Economists argue that an understanding of opportunity cost is crucial to the examination of choices.

As the set of available alternatives changes, we expect that the choices individuals make will change. A rainy day could change the opportunity cost of reading a good book; we might expect more reading to get done in bad than in good weather. A high income can make it very costly to take a day off; we might expect highly paid individuals to work more hours than those who are not paid as well. If individuals are maximizing their level of satisfaction and firms are maximizing profits, then a change in the set of alternatives they face may affect their choices in a predictable way.

The emphasis on opportunity costs is an emphasis on the examination of alternatives. One benefit of the economic way of thinking is that it pushes us to think about the value of alternatives in each problem involving choice.

Individuals Maximize in Pursuing Self-Interest

What motivates people as they make choices? Perhaps more than anything else, it is the economist's answer to this question that distinguishes economics from other fields.

Economists assume that individuals make choices that they expect will create the maximum value of some objective, given the constraints they face. Furthermore, economists assume that people's objectives will be those that serve their own self-interest.

Economists assume, for example, that the owners of business firms seek to maximize profit. Given the assumed goal of profit maximization, economists can predict how firms in an industry will respond to changes in the markets in which they operate. As labor costs in the United States rise, for example, economists are not surprised to see firms moving some of their manufacturing operations overseas.

Similarly, economists assume that maximizing behavior is at work when they examine the behavior of consumers. In studying consumers, economists assume that individual consumers make choices aimed at maximizing their level of satisfaction. In the next chapter, we will look at the results of the shift from skiing to snowboarding; that is a shift that reflects the pursuit of self-interest by consumers and by manufacturers.

In assuming that people pursue their self-interest, economists are not assuming people are selfish. People clearly gain satisfaction by helping others, as suggested by the large charitable contributions people make. Pursuing one's own self-interest means pursuing the things that give one satisfaction. It need not imply greed or selfishness.

Choices Are Made at the Margin

Economists argue that most choices are made "at the margin." The **margin** is the current level of an activity. Think of it as the edge from which a choice is to be made. A **choice at the margin** is a decision to do a little more or a little less of something.

margin

The current level of an activity.

choice at the margin

A decision to do a little more or a little less of something.

Assessing choices at the margin can lead to extremely useful insights. Consider, for example, the problem of curtailing water consumption when the amount of water available falls short of the amount people now use. Economists argue that one way to induce people to conserve water is to raise its price. A common response to this recommendation is that a higher price would have no effect on water consumption, because water is a necessity. Many people assert that prices do not affect water consumption because people "need" water.

But choices in water consumption, like virtually all choices, are made at the margin. Individuals do not make choices about whether they should or should not consume water. Rather, they decide whether to consume a little more or a little less water. Household water consumption in the United States totals about 105 gallons per person per day. Think of that starting point as the edge from which a choice at the margin in water consumption is made. Could a higher price cause you to use less water brushing your teeth, take shorter showers, or water your lawn less? Could a higher price cause people to reduce their use, say, to 104 gallons per person per day? To 103? When we examine the choice to consume water at the margin, the notion that a higher price would reduce consumption seems much more plausible. Prices affect our consumption of water because choices in water consumption, like other choices, are made at the margin.

The elements of opportunity cost, maximization, and choices at the margin can be found in each of two broad areas of economic analysis: microeconomics and macroeconomics. Your economics course, for example, may be designated as a "micro" or as a "macro" course. We will look at these two areas of economic thought in the next section.

2.2 Microeconomics and Macroeconomics

The field of economics is typically divided into two broad realms: microeconomics and macroeconomics. It is important to see the distinctions between these broad areas of study.

Microeconomics is the branch of economics that focuses on the choices made by individual decision-making units in the economy—typically consumers and firms—and the impacts those choices have on individual markets. **Macroeconomics** is the branch of economics that focuses on the impact of choices on the total, or aggregate, level of economic activity.

microeconomics

The branch of economics that focuses on the choices made by consumers and firms and the impacts those choices have on individual markets.

macroeconomics

The branch of economics that focuses on the impact of choices on the total, or aggregate, level of economic activity.

Why do tickets to the best concerts cost so much? How does the threat of global warming affect real estate prices in coastal areas? Why do women end up doing most of the housework? Why do senior citizens get discounts on public transit systems? These questions are generally regarded as microeconomic because they focus on individual units or markets in the economy.

Is the total level of economic activity rising or falling? Is the rate of inflation increasing or decreasing? What is happening to the unemployment rate? These are questions that deal with aggregates, or totals, in the economy; they are problems of macroeconomics. The question about the level of economic activity, for example, refers to the total value of all goods and services produced in the economy. Inflation is a measure of the rate of change in the average price level for the entire economy; it is a macroeconomic problem. The total levels of employment and unemployment in the economy represent the aggregate of all labor markets; unemployment is also a topic of macroeconomics.

Both microeconomics and macroeconomics give attention to individual markets. But in microeconomics that attention is an end in itself; in macroeconomics it is aimed at explaining the movement of major economic aggregates—the level of total output, the level of employment, and the price level.

We have now examined the characteristics that define the economic way of thinking and the two branches of this way of thinking: microeconomics and macroeconomics. In the next section, we will have a look at what one can do with training in economics.

2.3 Putting Economics to Work

Economics is one way of looking at the world. Because the economic way of thinking has proven quite useful, training in economics can be put to work in a wide range of fields. One, of course, is in work as an economist. Undergraduate work in economics can be applied to other careers as well.

Careers in Economics

Economists work in three types of organizations. About 58% of economists work for government agencies.[1] The remainder work for business firms or in colleges and universities.

Economists working for business firms and government agencies sometimes forecast economic activity to assist their employers in planning. They also apply economic analysis to the activities of the firms or agencies for which they work or consult. Economists employed at colleges and universities teach and conduct research.

Peruse the website of your college or university's economics department. Chances are the department will discuss the wide variety of occupations that their economics majors enter. Unlike engineering and accounting majors, economics and other social science majors tend to be distributed over a broad range of occupations.

Applying Economics to Other Fields

Suppose that you are considering something other than a career in economics. Would choosing to study economics help you?

The evidence suggests it may. Suppose, for example, that you are considering law school. The study of law requires keen analytical skills; studying economics sharpens such skills. Economists have traditionally argued that undergraduate work in economics serves as excellent preparation for law school. Economist Michael Nieswiadomy of the University of North Texas collected data on Law School Admittance Test (LSAT) scores for undergraduate majors listed by 2,200 or more students taking the test in 2003. Table 1.1 gives the scores, as well as the ranking for each of these majors, in 2003 and in two previous years in which the rankings were compiled. In rankings for all three years, economics majors recorded the highest scores.

TABLE 1.1 LSAT Scores and Undergraduate Majors

Here are the average LSAT scores and rankings for the 12 undergraduate majors with more than 2200 students taking the test to enter law school in the 2003–2004 academic year.

Major field	LSAT average 2003–2004	2003–2004 Rank	1994–1995 Rank	1991–1992 Rank
Economics	156.6	1	1	1
Engineering	155.4	2	4	2
History	155.0	3	2	3
English	154.3	4	3	4
Finance	152.6	5	6	5
Political science	152.1	6	9	9
Psychology	152.1	7	7	8
Accounting	151.1	8	8	6
Communications	150.5	9	10	10
Sociology	150.2	10	12	13
Bus. Administration	149.6	11	13	12
Criminal Justice	144.7	12	14	14

Source: Michael Nieswiadomy, "LSAT Scores of Economics Majors: 2003–2004 Class Update," Journal of Economic Education, 37(2) (Spring 2006): 244–247 and Michael Nieswiadomy, "LSAT Scores of Economics Majors" Journal of Economic Education, 29(4) (Fall 1998): 377–379.

Did the strong performance by economics, engineering, and history majors mean that training in those fields sharpens analytical skills tested in the LSAT, or that students with good analytical skills are more likely to major in them? Both factors were probably at work. Economics clearly attracts students with good analytical skills—and studying economics helps develop those skills.

Economics majors shine in other areas as well. According to the Bureau of Labor Statistics *Occupational Outlook Handbook*, a strong background in economic theory, mathematics, and statistics provides the basis for competing for the best job opportunities, particularly research assistant positions, in a broad range of fields. Many graduates with bachelor's degrees will find good jobs in industry and business as management or sales trainees or as administrative assistants. Because economists are concerned with understanding and interpreting financial matters, among other subjects, they will also be attracted to and qualified for jobs as financial managers, financial analysts, underwriters, actuaries, securities and financial services sales workers, credit analysts, loan and budget officers, and urban and regional planners.

Table 1.2 shows average yearly salary offers for bachelor degree candidates for May 2006 and the outlook for related occupations to 2014.

TABLE 1.2 Average Yearly Salary Offers, May 2006 and Occupational Outlook 2004–2014, Selected Majors/Occupations

Undergraduate major	Average $ Offer May, 2006	Projected % Change in Total Employment in Occupation 2004–2014
Computer Engineering	$54,200	10.1
Electrical/Electronic Engineering	54,053	11.8
Computer Science	50,892	25.6
Accounting	46,188	22.4
Economics and Finance	45,058	12.4
Management Information Systems	44,755	25.9
Logistics and Materials Management	43,426	13.2
Business Administration	40,976	17.0
Environmental Sciences (including forestry and conservation science)	39,750	6.3
Other Business Majors (e.g., Marketing)	37,446	20.8
Human Resources (incl. Labor Relations)	36,256	15.9
Geology and Geological Sciences	35,034	8.3
Sociology	33,752	4.7
Political Science/Government	33,151	7.3
Liberal Arts & Sciences (general studies)	32,627	na
Public Relations	32,623	21.7
Special Education	31,817	23.3
Elementary Education	31,778	18.2
Foreign Languages	31,364	na
Letters (incl. English)	31,204	20.4
Other Social Sciences (Including Criminal Justice and History)	30,788	12.3
Psychology	30,308	9.9
Pre-elementary Education	27,550	22.4
Social Work	25,865	19.6
Visual and Performing Arts	21,726	15.2

Sources: National Association of Colleges and Employers, Salary Survey, Spring 2006 http://naceweb.org; Bureau of Labor Statistics, 2006–2007 edition of the Occupational Outlook Handbook; Occupational Employment, Training, and Earnings: Educational Level Report (May, 2006) URL: http://data.bls.gov/oep/noeted/empoptd.jsp (note: na = not reported; that is, no specific occupation was reported in BLS report; Other business majors, Other social sciences, Social work (including Sociology), and Environmental Sciences are weighted averages of various disciplines, calculated by authors.)

One's choice of a major, or minor, is not likely to be based solely on considerations of potential earnings or the prospect of landing a spot in law school. You will also consider your interests and abilities in making a decision about whether to pursue further study in economics. And, of course, you will consider the expected benefits of alternative courses of study. What is *your* opportunity cost of pursuing study of economics? Does studying more economics serve your interests and will doing so maximize your satisfaction level? These considerations may be on your mind as you begin to study economics at the college level and obviously students will make many different choices. But, should you decide to pursue a major or minor in economics, you should know that a background in this field is likely to serve you well in a wide range of careers.

KEY TAKEAWAYS

- Economists focus on the opportunity costs of choices, they assume that individuals make choices in a way that maximizes the value of an objective defined in terms of their own self-interest, and they assume that individuals make those choices at the margin.
- Economics is divided into two broad areas: microeconomics and macroeconomics.
- A wide range of career opportunities is open to economics majors. Empirical evidence suggests that students who enter the job market with a major in economics tend to earn more than do students in most other majors. Further, economics majors do particularly well on the LSAT.

TRY IT!

The Department of Agriculture estimated that the expenditures a middle-income, husband–wife family of three would incur to raise one additional child from birth in 2005 to age 17 would be $250,530. In what way does this estimate illustrate the economic way of thinking? Would the Department's estimate be an example of microeconomic or of macroeconomic analysis? Why?

Case in Point: The Financial Payoff to Studying Economics

© 2010 Jupiterimages Corporation

College economics professors have long argued that studying economics is good preparation for a variety of careers. A recent study suggests they are right and that studying economics is even likely to make students more prosperous. Students who major in economics but did not pursue graduate work are likely to earn more than students in virtually every other college major. Students who major in economics and then go on to law school or an MBA program are likely to earn more than students who approach those areas of study having majored in most other areas.

Economists Dan A. Black, Seth Sanders, and Lowell Taylor used the 1993 National Survey of College Graduates, which included more than 86,000 college-educated workers between the ages of 25 and 55 that asked what field they had majored in. They then controlled for variables such as gender, race, and ethnicity. They found that students who had not done graduate work and had majored in economics earned more than students in any other major except engineering. Specifically, economics majors earned about 13% more than other social sciences majors, 11% more than business administration majors, and about the same as natural science and accounting majors. The economics majors in their survey, like those who majored in other social sciences and business administration and unlike those who majored in engineering or accounting, were spread out over a wide range of occupations but with many in management positions.

Based on the survey they used, over 40% of economics majors went on to earn graduate degrees, many in law and business. Economics majors ranked first in terms of wages, as compared to other law school graduates with the 12 most common pre-law majors (including such majors as business administration, finance, English, history, psychology, and political science). MBA graduates who had majored in economics earned more than those who had majored in any other field except chemical engineering. Specifically, undergraduate economics majors with MBAs earned about 15% more than those who had majored in other disciplines represented in the survey, including business-related majors.

It is remarkable that all of the business-related majors generated salaries much lower than those earned by economics majors with an MBA. One could argue that this reflects self-selection; that students who major in economics are simply brighter. But, students who major in physics have high SAT scores, yet they, too, earned wages that were about 20% lower than MBA students who had majored in economics. This finding lends some credence to the notion that the marketplace rewards training in the economic way of thinking.

Source: Dan A. Black, Seth Sanders, and Lowell Taylor, "The Economic Reward for Studying Economics," Economic Inquiry, 41(3), July 2003, 365–377.

ANSWER TO TRY IT! PROBLEM

The information given suggests one element of the economic way of thinking: assessing the choice at the margin. The estimate reflects the cost of one more child for a family that already has one. It is not clear from the information given how close the estimate of cost comes to the economic concept of opportunity cost. The Department of Agriculture's estimate included such costs as housing, food, transportation, clothing, health care, child care, and education. An economist would add the value of the best alternative use of the additional time that will be required for the child. If the couple is looking far ahead, it may want to consider the opportunity cost of sending a child to college. And, if it is looking *very* far ahead, it may want to consider the fact that nearly half of all parents over the age of 50 support at least one child over the age of 21. This is a problem in microeconomic analysis, because it focuses on the choices of individual households.

3. THE ECONOMISTS' TOOL KIT

LEARNING OBJECTIVES

1. **Explain how economists test hypotheses, develop economic theories, and use models in their analyses.**
2. **Explain how the all-other-things unchanged (ceteris paribus) problem and the fallacy of false cause affect the testing of economic hypotheses and how economists try to overcome these problems.**
3. **Distinguish between normative and positive statements.**

Economics differs from other social sciences because of its emphasis on opportunity cost, the assumption of maximization in terms of one's own self-interest, and the analysis of choices at the margin. But certainly much of the basic methodology of economics and many of its difficulties are common to every social science—indeed, to every science. This section explores the application of the scientific method to economics.

Researchers often examine relationships between variables. A **variable** is something whose value can change. By contrast, a **constant** is something whose value does not change. The speed at which a car is traveling is an example of a variable. The number of minutes in an hour is an example of a constant.

variable

Something whose value can change.

constant

Something whose value does not change.

scientific method

A systematic set of procedures through which knowledge is created.

hypothesis

An assertion of a relationship between two or more variables that could be proven to be false.

Research is generally conducted within a framework called the **scientific method**, a systematic set of procedures through which knowledge is created. In the scientific method, hypotheses are suggested and then tested. A **hypothesis** is an assertion of a relationship between two or more variables that could be proven to be false. A statement is not a hypothesis if no conceivable test could show it to be false. The statement "Plants like sunshine" is not a hypothesis; there is no way to test whether plants like sunshine or not, so it is impossible to prove the statement false. The statement "Increased solar radiation increases the rate of plant growth" is a hypothesis; experiments could be done to show the relationship between solar radiation and plant growth. If solar radiation were shown to be unrelated to plant growth or to retard plant growth, then the hypothesis would be demonstrated to be false.

If a test reveals that a particular hypothesis is false, then the hypothesis is rejected or modified. In the case of the hypothesis about solar radiation and plant growth, we would probably find that more sunlight increases plant growth over some range but that too much can actually retard plant growth. Such results would lead us to modify our hypothesis about the relationship between solar radiation and plant growth.

theory

A hypothesis that has not been rejected after widespread testing and that wins general acceptance.

law

A theory that has been subjected to even more testing and that has won virtually universal acceptance.

If the tests of a hypothesis yield results consistent with it, then further tests are conducted. A hypothesis that has not been rejected after widespread testing and that wins general acceptance is commonly called a **theory**. A theory that has been subjected to even more testing and that has won virtually universal acceptance becomes a **law**. We will examine two economic laws in the next two chapters.

Even a hypothesis that has achieved the status of a law cannot be proven true. There is always a possibility that someone may find a case that invalidates the hypothesis. That possibility means that nothing in economics, or in any other social science, or in any science, can ever be *proven* true. We can have great confidence in a particular proposition, but it is always a mistake to assert that it is "proven."

3.1 Models in Economics

model

A set of simplifying assumptions about some aspect of the real world.

All scientific thought involves simplifications of reality. The real world is far too complex for the human mind—or the most powerful computer—to consider. Scientists use models instead. A **model** is a set of simplifying assumptions about some aspect of the real world. Models are always based on assumed conditions that are simpler than those of the real world, assumptions that are necessarily false. A model of the real world cannot *be* the real world.

We will encounter our first economic model in Chapter 35. For that model, we will assume that an economy can produce only two goods. Then we will explore the model of demand and supply. One of the assumptions we will make there is that all the goods produced by firms in a particular market are identical. Of course, real economies and real markets are not that simple. Reality is never as simple as a model; one point of a model is to simplify the world to improve our understanding of it.

Economists often use graphs to represent economic models. The appendix to this chapter provides a quick, refresher course, if you think you need one, on understanding, building, and using graphs.

Models in economics also help us to generate hypotheses about the real world. In the next section, we will examine some of the problems we encounter in testing those hypotheses.

3.2 Testing Hypotheses in Economics

Here is a hypothesis suggested by the model of demand and supply: an increase in the price of gasoline will reduce the quantity of gasoline consumers demand. How might we test such a hypothesis?

Economists try to test hypotheses such as this one by observing actual behavior and using empirical (that is, real-world) data. The average retail price of gasoline in the United States rose from an average of $2.12 per gallon on May 22, 2005 to $2.88 per gallon on May 22, 2006. The number of gallons of gasoline consumed by U.S. motorists rose 0.3% during that period.

The small increase in the quantity of gasoline consumed by motorists as its price rose is inconsistent with the hypothesis that an increased price will lead to an reduction in the quantity demanded. Does that mean that we should dismiss the original hypothesis? On the contrary, we must be cautious in assessing this evidence. Several problems exist in interpreting any set of economic data. One problem is that several things may be changing at once; another is that the initial event may be unrelated to the event that follows. The next two sections examine these problems in detail.

The All-Other-Things-Unchanged Problem

ceteris paribus

A Latin phrase that means, "all other things unchanged."

The hypothesis that an increase in the price of gasoline produces a reduction in the quantity demanded by consumers carries with it the assumption that there are no other changes that might also affect consumer demand. A better statement of the hypothesis would be: An increase in the price of gasoline will reduce the quantity consumers demand, ceteris paribus. **Ceteris paribus** is a Latin phrase that means "all other things unchanged."

But things changed between May 2005 and May 2006. Economic activity and incomes rose both in the United States and in many other countries, particularly China, and people with higher incomes are likely to buy more gasoline. Employment rose as well, and people with jobs use more gasoline as they drive to work. Population in the United States grew during the period. In short, many things happened during the period, all of which tended to increase the quantity of gasoline people purchased.

Our observation of the gasoline market between May 2005 and May 2006 did not offer a conclusive test of the hypothesis that an increase in the price of gasoline would lead to a reduction in the quantity demanded by consumers. Other things changed and affected gasoline consumption. Such problems are likely to affect any analysis of economic events. We cannot ask the world to stand still while we conduct experiments in economic phenomena. Economists employ a variety of statistical methods to allow them to isolate the impact of single events such as price changes, but they can never be certain that they have accurately isolated the impact of a single event in a world in which virtually everything is changing all the time.

In laboratory sciences such as chemistry and biology, it is relatively easy to conduct experiments in which only selected things change and all other factors are held constant. The economists' laboratory is the real world; thus, economists do not generally have the luxury of conducting controlled experiments.

The Fallacy of False Cause

Hypotheses in economics typically specify a relationship in which a change in one variable causes another to change. We call the variable that responds to the change the **dependent variable**; the variable that induces a change is called the **independent variable**. Sometimes the fact that two variables move together can suggest the false conclusion that one of the variables has acted as an independent variable that has caused the change we observe in the dependent variable.

dependent variable

A variable that responds to change.

independent variable

A variable that induces a change.

Consider the following hypothesis: People wearing shorts cause warm weather. Certainly, we observe that more people wear shorts when the weather is warm. Presumably, though, it is the warm weather that causes people to wear shorts rather than the wearing of shorts that causes warm weather; it would be incorrect to infer from this that people cause warm weather by wearing shorts.

Reaching the incorrect conclusion that one event causes another because the two events tend to occur together is called the **fallacy of false cause**. The accompanying essay on baldness and heart disease suggests an example of this fallacy.

fallacy of false cause

The incorrect assumption that one event causes another because the two events tend to occur together.

Because of the danger of the fallacy of false cause, economists use special statistical tests that are designed to determine whether changes in one thing actually do cause changes observed in another. Given the inability to perform controlled experiments, however, these tests do not always offer convincing evidence that persuades all economists that one thing does, in fact, cause changes in another.

In the case of gasoline prices and consumption between May 2005 and May 2006, there is good theoretical reason to believe the price increase should lead to a reduction in the quantity consumers demand. And economists have tested the hypothesis about price and the quantity demanded quite extensively. They have developed elaborate statistical tests aimed at ruling out problems of the fallacy of false cause. While we cannot prove that an increase in price will, ceteris paribus, lead to a reduction in the quantity consumers demand, we can have considerable confidence in the proposition.

Normative and Positive Statements

Two kinds of assertions in economics can be subjected to testing. We have already examined one, the hypothesis. Another testable assertion is a statement of fact, such as "It is raining outside" or "Microsoft is the largest producer of operating systems for personal computers in the world." Like hypotheses, such assertions can be demonstrated to be false. Unlike hypotheses, they can also be shown to be correct. A statement of fact or a hypothesis is a **positive statement**.

positive statement

A statement of fact or a hypothesis.

Although people often disagree about positive statements, such disagreements can ultimately be resolved through investigation. There is another category of assertions, however, for which investigation can never resolve differences. A **normative statement** is one that makes a value judgment. Such a judgment is the opinion of the speaker; no one can "prove" that the statement is or is not correct. Here are some examples of normative statements in economics: "We ought to do more to help the poor." "People in the United States should save more." "Corporate profits are too high." The statements are based on the values of the person who makes them. They cannot be proven false.

normative statement

A statement that makes a value judgment.

Because people have different values, normative statements often provoke disagreement. An economist whose values lead him or her to conclude that we should provide more help for the poor will disagree with one whose values lead to a conclusion that we should not. Because no test exists for these values, these two economists will continue to disagree, unless one persuades the other to adopt a different set of values. Many of the disagreements among economists are based on such differences in values and therefore are unlikely to be resolved.

KEY TAKEAWAYS

- Economists try to employ the scientific method in their research.
- Scientists cannot prove a hypothesis to be true; they can only fail to prove it false.
- Economists, like other social scientists and scientists, use models to assist them in their analyses.
- Two problems inherent in tests of hypotheses in economics are the all-other-things-unchanged problem and the fallacy of false cause.
- Positive statements are factual and can be tested. Normative statements are value judgments that cannot be tested. Many of the disagreements among economists stem from differences in values.

TRY IT!

Look again at the data in Table 1.1. Now consider the hypothesis: "Majoring in economics will result in a higher LSAT score." Are the data given consistent with this hypothesis? Do the data prove that this hypothesis is correct? What fallacy might be involved in accepting the hypothesis?

Case in Point: Does Baldness Cause Heart Disease?

© 2010 Jupiterimages Corporation

A website called embarrassingproblems.com received the following email:

> *"Dear Dr. Margaret,*
>
> *"I seem to be going bald. According to your website, this means I'm more likely to have a heart attack. If I take a drug to prevent hair loss, will it reduce my risk of a heart attack?"*

What did Dr. Margaret answer? Most importantly, she did not recommend that the questioner take drugs to treat his baldness, because doctors do not think that the baldness causes the heart disease. A more likely explanation for the association between baldness and heart disease is that both conditions are affected by an underlying factor. While noting that more research needs to be done, one hypothesis that Dr. Margaret offers is that higher testosterone levels might be triggering both the hair loss and the heart disease. The good news for people with early balding (which is really where the association with increased risk of heart disease has been observed) is that they have a signal that might lead them to be checked early on for heart disease.

Source: http://www.embarrassingproblems.com/problems/problempage230701.htm.

ANSWER TO TRY IT! PROBLEM

The data are consistent with the hypothesis, but it is never possible to prove that a hypothesis is correct. Accepting the hypothesis could involve the fallacy of false cause; students who major in economics may already have the analytical skills needed to do well on the exam.

4. REVIEW AND PRACTICE

Summary

Choices are forced on us by scarcity; economists study the choices that people make. Scarce goods are those for which the choice of one alternative requires giving up another. The opportunity cost of any choice is the value of the best alternative forgone in making that choice.

Some key choices assessed by economists include what to produce, how to produce it, and for whom it should be produced. Economics is distinguished from other academic disciplines that also study choices by an emphasis on the central importance of opportunity costs in evaluating choices, the assumption of maximizing behavior that serves the interests of individual decision makers, and a focus on evaluating choices at the margin.

Economic analyses may be aimed at explaining individual choice or choices in an individual market; such investigations are largely the focus of microeconomics. The analysis of the impact of those individual choices on such aggregates as total output, the level of employment, and the price level is the concern of macroeconomics.

Working within the framework of the scientific method, economists formulate hypotheses and then test them. These tests can only refute a hypothesis; hypotheses in science cannot be proved. A hypothesis that has been widely tested often comes to be regarded as a theory; one that has won virtually universal acceptance is a law. Because of the complexity of the real world, economists rely on models that rest on a series of simplifying assumptions. The models are used to generate hypotheses about the economy that can be tested using real-world data.

Statements of fact and hypotheses are positive statements. Normative statements, unlike positive statements, cannot be tested and provide a source for potential disagreement.

PROBLEMS

1. Why does the fact that something is scarce require that we make choices?
2. Does the fact that something is abundant mean it is not scarce in the economic sense? Why or why not?
3. In some countries, such as Cuba and North Korea, the government makes most of the decisions about what will be produced, how it will be produced, and for whom. Does the fact that these choices are made by the government eliminate scarcity in these countries? Why or why not?
4. Explain what is meant by the opportunity cost of a choice.
5. What is the approximate dollar cost of the tuition and other fees associated with the economics course you are taking? Does this dollar cost fully reflect the opportunity cost to you of taking the course?
6. In the Case in Point essay "The Rising Cost of Energy," what would be some of the things that would be included in an estimate of the opportunity cost of preserving part of northern Alberta Canada by prohibiting heavy crude oil extraction? Do you think that the increased extraction represents the best use of the land? Why or why not?
7. Indicate whether each of the following is a topic of microeconomics or macroeconomics:
    a. The impact of higher oil prices on the production of steel
    b. The increased demand in the last 15 years for exotic dietary supplements
    c. The surge in aggregate economic activity that hit much of Asia late in the early 2000s
    d. The sharp increases in U.S. employment and total output that occurred between 2003 and 2007
    e. The impact of preservation of wilderness areas on the logging industry and on the price of lumber
8. Determine whether each of the following raises a "what," "how," or "for whom" issue. Are the statements normative or positive?
    a. A requirement that aluminum used in cars be made from recycled materials will raise the price of automobiles.
    b. The federal government does not spend enough for children.
    c. An increase in police resources provided to the inner city will lower the crime rate.
    d. Automation destroys jobs.
    e. Efforts to improve the environment tend to reduce production and employment.
    f. Japanese firms should be more willing to hire additional workers when production rises and to lay off workers when production falls.
    g. Access to health care should not be limited by income.
9. Your time is a scarce resource. What if the quantity of time were increased, say to 48 hours per day, and everyone still lived as many days as before. Would time still be scarce?
10. Most college students are under age 25. Give two explanations for this—one based on the benefits people of different ages are likely to receive from higher education and one based on the opportunity costs of a college education to students of different ages.
11. Some municipal water companies charge customers a flat fee each month, regardless of the amount of water they consume. Others meter water use and charge according to the quantity of water customers use. Compare the way the two systems affect the cost of water use at the margin.
12. How might you test each of the following hypotheses? Suggest some problems that might arise in each test due to the ceteris paribus (all-other-things-unchanged) problem and the fallacy of false cause.
    a. Reducing the quantity of heroin available will increase total spending on heroin and increase the crime rate.
    b. Higher incomes make people happier.
    c. Higher incomes make people live longer.
13. Many models in physics and in chemistry assume the existence of a perfect vacuum (that is, a space entirely empty of matter). Yet we know that a perfect vacuum cannot exist. Are such models valid? Why are models based on assumptions that are essentially incorrect?
14. Suppose you were asked to test the proposition that publishing students' teacher evaluations causes grade inflation. What evidence might you want to consider? How would the inability to carry out controlled experiments make your analysis more difficult?
15. Referring to the Case in Point "Baldness and Heart Disease," explain the possible fallacy of false cause in concluding that baldness makes a person more likely to have heart disease.
16. In 2005 the Food and Drug Administration ordered that Vioxx and other popular drugs for treating the pain of arthritis be withdrawn from the market. The order resulted from a finding that people taking the drugs had an increased risk of cardiovascular problems. Some researchers criticized the government's action, arguing that concluding that the drugs caused the cardiovascular problems represented an example of the fallacy of false cause. Can you think of any reason why this might be the case?

ENDNOTES

1. Bureau of Labor Statistics *Occupational Outlook* at http://www.bls.gov/oco/.

CHAPTER 2

Confronting Scarcity: Choices in Production

START UP: TIGHTENING SECURITY AT THE WORLD'S AIRPORTS

Do you want safer air travel or not? While that question is seldom asked so bluntly, any person who travels by air can tell you that our collective answer has been "yes," and it has been accompanied by increases in security and its associated costs at airports all over the world. Why? In short, "9/11." Terrorists hijacked four U.S. commercial airliners on September 11, 2001, and the tragic results that followed led to a sharp tightening in airport security.

In an effort to prevent similar disasters, airport security officials scrutinize luggage and passengers more carefully than ever before. In the months following 9/11, delays of as much as three hours were common as agents tried to assure that no weapons or bombs could be smuggled onto another plane.

"What to produce?" is a fundamental economic question. Every economy must answer this question. Should it produce more education, better health care, improved transportation, a cleaner environment? There are limits to what a nation can produce; deciding to produce more of one thing inevitably means producing less of something else. Individuals in much of the world, after the tragedy of 9/11, clearly were willing to give up time, and a fair amount of individual privacy, in an effort to obtain greater security. Nations and individual cities also devoted additional resources to police and other forms of protection in an effort to prevent tragedies such as 9/11. People all over the world chose to produce less of other goods in order to devote more resources to the production of greater security. And, as of early 2009, the choice to devote more resources to security had paid off; there had been no similar hijackings in the United States.

In this chapter we use our first model, the production possibilities model, to examine the nature of choices to produce more of some goods and less of others. As its name suggests, the **production possibilities model** shows the goods and services that an economy is capable of producing—its possibilities—given the factors of production and the technology it has available. The model specifies what it means to use resources fully and efficiently and suggests some important implications for international trade. We can also use the model to illustrate economic growth, a process that expands the set of production possibilities available to an economy.

production possibilities model

Model that shows the goods and services that an economy is capable of producing—its possibilities—given the factors of production and the technology it has available.

We then turn to an examination of the type of economic system in which choices are made. An **economic system** is the set of rules that define how an economy's resources are to be owned and how decisions about their use are to be made. We will see that economic systems differ in terms of how they answer the fundamental economic questions. Many of the world's economic systems, including the systems that prevail in North America, Europe, and much of Asia and Central and South America, rely on individuals operating in a market economy to make those choices. Other economic systems, including those of Cuba and North Korea today and historically

economic system

The set of rules that define how an economy's resources are to be owned and how decisions about their use are to be made.

those of the former Soviet Union, Soviet bloc countries, and China, rely—or relied—on government to make these choices. Different economic systems result in different sets of choices and thus different outcomes; the fact that market economies generally outperform the others when it comes to providing more of the things that people want helps to explain the dramatic shift from government-dominated toward market-dominated economic systems that has occurred throughout the world in the past 25 years. The chapter concludes with an examination of the role of government in an economy that relies chiefly on markets to allocate goods and services.

1. FACTORS OF PRODUCTION

LEARNING OBJECTIVES

1. **Define the three factors of production—labor, capital, and natural resources.**
2. **Explain the role of technology and entrepreneurs in the utilization of the economy's factors of production.**

Choices concerning what goods and services to produce are choices about an economy's use of its **factors of production**, the resources available to it for the production of goods and services. The value, or satisfaction, that people derive from the goods and services they consume and the activities they pursue is called **utility**. Ultimately, then, an economy's factors of production create utility; they serve the interests of people.

The factors of production in an economy are its labor, capital, and natural resources. **Labor** is the human effort that can be applied to the production of goods and services. People who are employed or would like to be are considered part of the labor available to the economy. **Capital** is a factor of production that has been produced for use in the production of other goods and services. Office buildings, machinery, and tools are examples of capital. **Natural resources** are the resources of nature that can be used for the production of goods and services.

In the next three sections, we will take a closer look at the factors of production we use to produce the goods and services we consume. The three basic building blocks of labor, capital, and natural resources may be used in different ways to produce different goods and services, but they still lie at the core of production. We will then look at the roles played by technology and entrepreneurs in putting these factors of production to work. As economists began to grapple with the problems of scarcity, choice, and opportunity cost two centuries ago, they focused on these concepts, just as they are likely to do two centuries hence.

factors of production

The resources available to the economy for the production of goods and services.

utility

The value, or satisfaction, that people derive from the goods and services they consume and the activities they pursue.

labor

The human effort that can be applied to the production of goods and services.

capital

A factor of production that has been produced for use in the production of other goods and services.

natural resources

The resources of nature that can be used for the production of goods and services.

1.1 Labor

Labor is human effort that can be applied to production. People who work to repair tires, pilot airplanes, teach children, or enforce laws are all part of the economy's labor. People who would like to work but have not found employment—who are unemployed—are also considered part of the labor available to the economy.

In some contexts, it is useful to distinguish two forms of labor. The first is the human equivalent of a natural resource. It is the natural ability an untrained, uneducated person brings to a particular production process. But most workers bring far more. The skills a worker has as a result of education, training, or experience that can be used in production are called **human capital**. Students who are attending a college or university are acquiring human capital. Workers who are gaining skills through experience or through training are acquiring human capital. Children who are learning to read are acquiring human capital.

human capital

The skills a worker has as a result of education, training, or experience that can be used in production.

The amount of labor available to an economy can be increased in two ways. One is to increase the total quantity of labor, either by increasing the number of people available to work or by increasing the average number of hours of work per week. The other is to increase the amount of human capital possessed by workers.

1.2 Capital

Long ago, when the first human beings walked the earth, they produced food by picking leaves or fruit off a plant or by catching an animal and eating it. We know that very early on, however, they began shaping stones into tools, apparently for use in butchering animals. Those tools were the first capital because they were produced for use in producing other goods—food and clothing.

Modern versions of the first stone tools include saws, meat cleavers, hooks, and grinders; all are used in butchering animals. Tools such as hammers, screwdrivers, and wrenches are also capital. Transportation equipment, such as cars and trucks, is capital. Facilities such as roads, bridges, ports, and airports are capital. Buildings, too, are capital; they help us to produce goods and services.

Capital does not consist solely of physical objects. The score for a new symphony is capital because it will be used to produce concerts. Computer software used by business firms or government agencies to produce goods and services is capital. Capital may thus include physical goods and intellectual discoveries. Any resource is capital if it satisfies two criteria:

1. The resource must have been produced.
2. The resource can be used to produce other goods and services.

One thing that is not considered capital is money. A firm cannot use money directly to produce other goods, so money does not satisfy the second criterion for capital. Firms can, however, use money to acquire capital. Money is a form of financial capital. **Financial capital** includes money and other "paper" assets (such as stocks and bonds) that represent claims on future payments. These financial assets are not capital, but they can be used directly or indirectly to purchase factors of production or goods and services.

financial capital

Includes money and other "paper" assets (such as stocks and bonds) that represent claims on future payments.

1.3 Natural Resources

There are two essential characteristics of natural resources. The first is that they are found in nature—that no human effort has been used to make or alter them. The second is that they can be used for the production of goods and services. That requires knowledge; we must know how to use the things we find in nature before they become resources.

Consider oil. Oil in the ground is a natural resource because it is found (not manufactured) and can be used to produce goods and services. However, 250 years ago oil was a nuisance, not a natural resource. Pennsylvania farmers in the eighteenth century who found oil oozing up through their soil were dismayed, not delighted. No one knew what could be done with the oil. It was not until the mid-nineteenth century that a method was found for refining oil into kerosene that could be used to generate energy, transforming oil into a natural resource. Oil is now used to make all sorts of things, including clothing, drugs, gasoline, and plastic. It became a natural resource because people discovered and implemented a way to use it.

Defining something as a natural resource only if it can be used to produce goods and services does not mean that a tree has value only for its wood or that a mountain has value only for its minerals. If people gain utility from the existence of a beautiful wilderness area, then that wilderness provides a service. The wilderness is thus a natural resource.

The natural resources available to us can be expanded in three ways. One is the discovery of new natural resources, such as the discovery of a deposit of ore containing titanium. The second is the discovery of new uses for resources, as happened when new techniques allowed oil to be put to productive use or sand to be used in manufacturing computer chips. The third is the discovery of new ways to extract natural resources in order to use them. New methods of discovering and mapping oil deposits have increased the world's supply of this important natural resource.

1.4 Technology and the Entrepreneur

technology

The knowledge that can be applied to the production of goods and services.

entrepreneur

A person who, operating within the context of a market economy, seeks to earn profits by finding new ways to organize factors of production.

Goods and services are produced using the factors of production available to the economy. Two things play a crucial role in putting these factors of production to work. The first is **technology**, the knowledge that can be applied to the production of goods and services. The second is an individual who plays a key role in a market economy: the entrepreneur. An **entrepreneur** is a person who, operating within the context of a market economy, seeks to earn profits by finding new ways to organize factors of production. In non-market economies the role of the entrepreneur is played by bureaucrats and other decision makers who respond to incentives other than profit to guide their choices about resource allocation decisions.

The interplay of entrepreneurs and technology affects all our lives. Entrepreneurs put new technologies to work every day, changing the way factors of production are used. Farmers and factory workers, engineers and electricians, technicians and teachers all work differently than they did just a few years ago, using new technologies introduced by entrepreneurs. The music you enjoy, the books you read, the athletic equipment with which you play are produced differently than they were five years ago. The book you are reading was written and manufactured using technologies that did not exist ten years ago. We can dispute whether all the changes have made our lives better. What we cannot dispute is that they have made our lives different.

KEY TAKEAWAYS

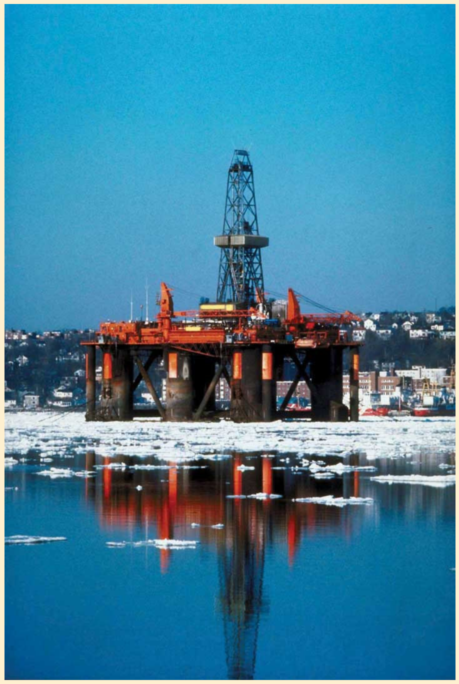
- Factors of production are the resources the economy has available to produce goods and services.
- Labor is the human effort that can be applied to the production of goods and services. Labor's contribution to an economy's output of goods and services can be increased either by increasing the quantity of labor or by increasing human capital.
- Capital is a factor of production that has been produced for use in the production of other goods and services.
- Natural resources are those things found in nature that can be used for the production of goods and services.
- Two keys to the utilization of an economy's factors of production are technology and, in the case of a market economic system, the efforts of entrepreneurs.

TRY IT!

Explain whether each of the following is labor, capital, or a natural resource.

1. An unemployed factory worker
2. A college professor
3. The library building on your campus
4. Yellowstone National Park
5. An untapped deposit of natural gas
6. The White House
7. The local power plant

Case in Point: Technology Cuts Costs, Boosts Productivity and Profits



Technology can seem an abstract force in the economy—important, but invisible.

It is not invisible to the 130 people who work on a Shell Oil Company oil rig called Mars, located in the deep waters of the Gulf of Mexico, about 160 miles southwest of Pensacola, Florida. The name Mars reflects its otherworld appearance—it extends 300 feet above the water's surface and has steel tendons that reach 3,000 feet to the floor of the gulf. This facility would not exist if it were not for the development of better oil discovery methods that include three-dimensional seismic mapping techniques, satellites that locate oil from space, and drills that can make turns as drilling foremen steer them by monitoring them on computer screens from the comfort of Mars. "We don't hit as many dry holes," commented Shell manager Miles Barrett. As a result of these new technologies, over the past two decades, the cost of discovering a barrel of oil dropped from $20 to under $5. And the technologies continue to improve. Three-dimensional surveys are being replaced with four-dimensional ones that allow geologists to see how the oil fields change over time.

The Mars project was destroyed by Hurricane Katrina in 2005. Royal Dutch Shell completed repairs in 2006—at a cost of $200 million. But, the facility is again pumping 130,000 barrels of oil per day and 150 million cubic feet of natural gas—the energy equivalent of an additional 26,000 barrels of oil.

Technology is doing more than helping energy companies track oil deposits. It is changing the way soft drinks and other grocery items are delivered to retail stores. For example, when a PepsiCo delivery driver arrives at a 7-Eleven, the driver keys into a handheld computer the inventory of soft drinks, chips, and other PepsiCo products. The information is transmitted to a main computer at the warehouse that begins processing the next order for that store. The result is that the driver can visit more stores in a day and PepsiCo can cover a given territory with fewer drivers and trucks.

New technology is even helping to produce more milk from cows. Ed Larsen, who owns a 1,200-cow dairy farm in Wisconsin, never gets up before dawn to milk the cows, the way he did as a boy. Rather, the cows are hooked up to electronic milkers. Computers measure each cow's output, and cows producing little milk are sent to a "hospital wing" for treatment. With the help of such technology, as well as better feed, today's dairy cows produce 50% more milk than did cows 20 years ago. Even though the number of dairy cows in the United States in the last 20 years has fallen 17%, milk output has increased 25%.

Who benefits from technological progress? Consumers gain from lower prices and better service. Workers gain: Their greater ability to produce goods and services translates into higher wages. And firms gain: Lower production costs mean higher profits. Of course, some people lose as technology advances. Some jobs are eliminated, and some firms find their services are no longer needed. One can argue about whether particular technological changes have improved our lives, but they have clearly made—and will continue to make—them far different.

Sources: David Ballingrud, "Drilling in the Gulf: Life on Mars," St. Petersburg Times (Florida), August 5, 2001, p. 1A; Barbara Hagenbaugh, "Dairy Farms Evolve to Survive," USA Today, August 7, 2003, p. 1B; Del Jones and Barbara Hansen, "Special Report: A Who's Who of Productivity," USA Today, August 30, 2001, p. 1B; and Christopher Helman, Shell Shocked, Forbes Online, July 27, 2006.

ANSWERS TO TRY IT! PROBLEMS

1. An unemployed factory worker could be put to work; he or she counts as labor.
2. A college professor is labor.
3. The library building on your campus is part of capital.
4. Yellowstone National Park. Those areas of the park left in their natural state are a natural resource. Facilities such as visitors' centers, roads, and campgrounds are capital.
5. An untapped deposit of natural gas is a natural resource. Once extracted and put in a storage tank, natural gas is capital.
6. The White House is capital.
7. The local power plant is capital.

2. THE PRODUCTION POSSIBILITIES CURVE

LEARNING OBJECTIVES

1. **Explain the concept of the production possibilities curve and understand the implications of its downward slope and bowed-out shape.**
2. **Use the production possibilities model to distinguish between full employment and situations of idle factors of production and between efficient and inefficient production.**
3. **Understand specialization and its relationship to the production possibilities model and comparative advantage.**

production possibilities curve

A graphical representation of the alternative combinations of goods and services an economy can produce.

An economy's factors of production are scarce; they cannot produce an unlimited quantity of goods and services. A **production possibilities curve** is a graphical representation of the alternative combinations of goods and services an economy can produce. It illustrates the production possibilities model. In drawing the production possibilities curve, we shall assume that the economy can produce only two goods and that the quantities of factors of production and the technology available to the economy are fixed.

2.1 Constructing a Production Possibilities Curve

To construct a production possibilities curve, we will begin with the case of a hypothetical firm, Alpine Sports, Inc., a specialized sports equipment manufacturer. Christie Ryder began the business 15 years ago with a single ski production facility near Killington ski resort in central Vermont. Ski sales grew, and she also saw demand for snowboards rising—particularly after snowboard competition events were included in the 2002 Winter Olympics in Salt Lake City. She added a second plant in a nearby town. The second plant, while smaller than the first, was designed to produce snowboards as well as skis. She also modified the first plant so that it could produce both snowboards and skis. Two years later she added a third plant in another town. While even smaller than the second plant, the third was primarily designed for snowboard production but could also produce skis.

We can think of each of Ms. Ryder's three plants as a miniature economy and analyze them using the production possibilities model. We assume that the factors of production and technology available to each of the plants operated by Alpine Sports are unchanged.

Suppose the first plant, Plant 1, can produce 200 pairs of skis per month when it produces only skis. When devoted solely to snowboards, it produces 100 snowboards per month. It can produce skis and snowboards simultaneously as well.

The table in Figure 2.2 gives three combinations of skis and snowboards that Plant 1 can produce each month. Combination A involves devoting the plant entirely to ski production; combination C means shifting all of the plant's resources to snowboard production; combination B involves the production of both goods. These values are plotted in a production possibilities curve for Plant 1. The curve is a downward-sloping straight line, indicating that there is a linear, negative relationship between the production of the two goods.

Neither skis nor snowboards is an independent or a dependent variable in the production possibilities model; we can assign either one to the vertical or to the horizontal axis. Here, we have placed the number of pairs of skis produced per month on the vertical axis and the number of snowboards produced per month on the horizontal axis.

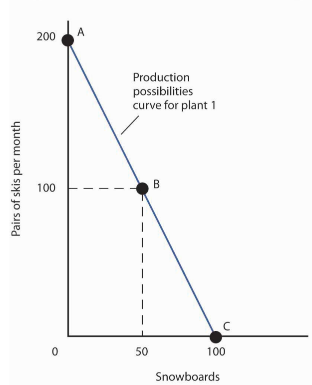
The negative slope of the production possibilities curve reflects the scarcity of the plant's capital and labor. Producing more snowboards requires shifting resources out of ski production and thus producing fewer skis. Producing more skis requires shifting resources out of snowboard production and thus producing fewer snowboards.

The slope of Plant 1's production possibilities curve measures the rate at which Alpine Sports must give up ski production to produce additional snowboards. Because the production possibilities curve for Plant 1 is linear, we can compute the slope between any two points on the curve and get the same result. Between points A and B, for example, the slope equals −2 pairs of skis/snowboard (equals −100 pairs of skis/50 snowboards). (Many students are helped when told to read this result as "−2 pairs of skis *per* snowboard.") We get the same value between points B and C, and between points A and C.

FIGURE 2.2 A Production Possibilities Curve

The table shows the combinations of pairs of skis and snowboards that Plant 1 is capable of producing each month. These are also illustrated with a production possibilities curve. Notice that this curve is linear.

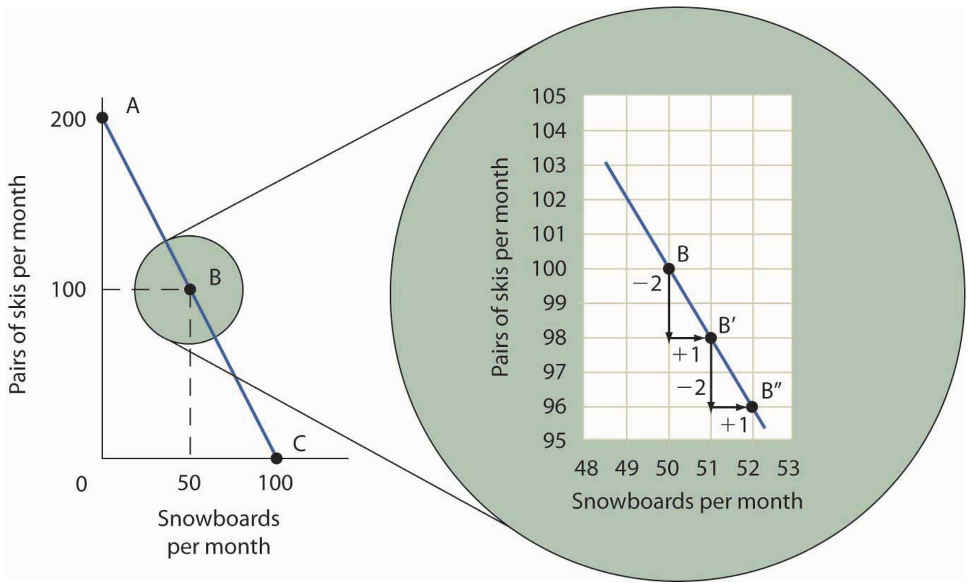
	Pairs of skis per month	Snowboards per month
A	200	0
B	100	50
C	0	100



To see this relationship more clearly, examine Figure 2.3. Suppose Plant 1 is producing 100 pairs of skis and 50 snowboards per month at point B. Now consider what would happen if Ms. Ryder decided to produce 1 more snowboard per month. The segment of the curve around point B is magnified in Figure 2.3. The slope between points B and B′ is −2 pairs of skis/snowboard. Producing 1 additional snowboard at point B′ requires giving up 2 pairs of skis. We can think of this as the opportunity cost of producing an additional snowboard at Plant 1. This opportunity cost equals the absolute value of the slope of the production possibilities curve.

FIGURE 2.3 The Slope of a Production Possibilities Curve

The slope of the linear production possibilities curve in Figure 2.2 is constant; it is −2 pairs of skis/snowboard. In the section of the curve shown here, the slope can be calculated between points B and B′. Expanding snowboard production to 51 snowboards per month from 50 snowboards per month requires a reduction in ski production to 98 pairs of skis per month from 100 pairs. The slope equals −2 pairs of skis/snowboard (that is, it must give up two pairs of skis to free up the resources necessary to produce one additional snowboard). To shift from B′ to B″, Alpine Sports must give up two more pairs of skis per snowboard. The absolute value of the slope of a production possibilities curve measures the opportunity cost of an additional unit of the good on the horizontal axis measured in terms of the quantity of the good on the vertical axis that must be forgone.

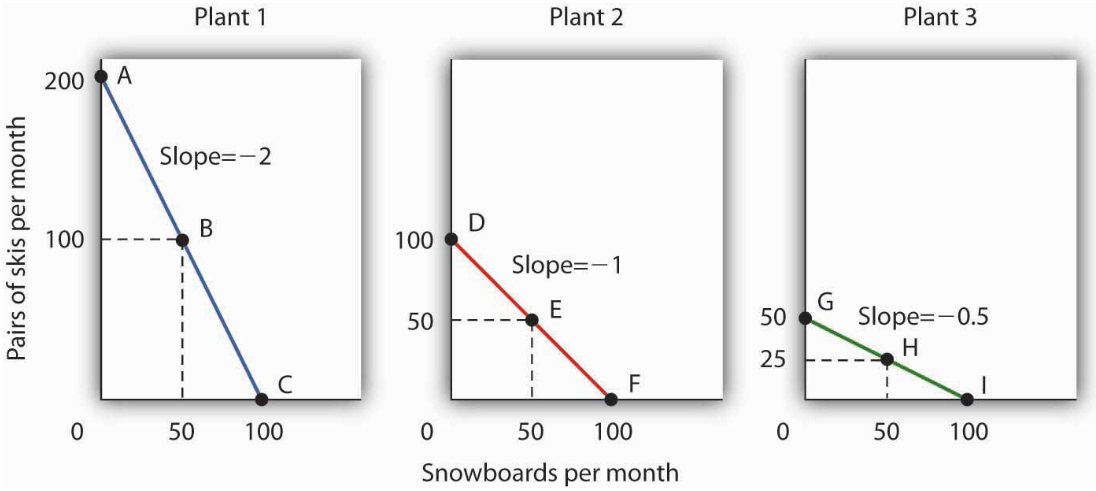


The absolute value of the slope of any production possibilities curve equals the opportunity cost of an additional unit of the good on the horizontal axis. It is the amount of the good on the vertical axis that must be given up in order to free up the resources required to produce one more unit of the good on the horizontal axis. We will make use of this important fact as we continue our investigation of the production possibilities curve.

Figure 2.4 shows production possibilities curves for each of the firm's three plants. Each of the plants, if devoted entirely to snowboards, could produce 100 snowboards. Plants 2 and 3, if devoted exclusively to ski production, can produce 100 and 50 pairs of skis per month, respectively. The exhibit gives the slopes of the production possibilities curves for each plant. The opportunity cost of an additional snowboard at each plant equals the absolute values of these slopes (that is, the number of pairs of skis that must be given up per snowboard).

FIGURE 2.4 Production Possibilities at Three Plants

The slopes of the production possibilities curves for each plant differ. The steeper the curve, the greater the opportunity cost of an additional snowboard. Here, the opportunity cost is lowest at Plant 3 and greatest at Plant 1.



The exhibit gives the slopes of the production possibilities curves for each of the firm's three plants. The opportunity cost of an additional snowboard at each plant equals the absolute values of these slopes. More generally, the absolute value of the slope of any production possibilities curve at any point gives the opportunity cost of an additional unit of the good on the horizontal axis, measured in terms of the number of units of the good on the vertical axis that must be forgone.

The greater the absolute value of the slope of the production possibilities curve, the greater the opportunity cost will be. The plant for which the opportunity cost of an additional snowboard is greatest is the plant with the steepest production possibilities curve; the plant for which the opportunity cost is lowest is the plant with the flattest production possibilities curve. The plant with the lowest opportunity cost of producing snowboards is Plant 3; its slope of −0.5 means that Ms. Ryder must give up half a pair of skis in that plant to produce an additional snowboard. In Plant 2, she must give up one pair of skis to gain one more snowboard. We have already seen that an additional snowboard requires giving up two pairs of skis in Plant 1.

2.2 Comparative Advantage and the Production Possibilities Curve

To construct a combined production possibilities curve for all three plants, we can begin by asking how many pairs of skis Alpine Sports could produce if it were producing only skis. To find this quantity, we add up the values at the vertical intercepts of each of the production possibilities curves in Figure 2.4. These intercepts tell us the maximum number of pairs of skis each plant can produce. Plant 1 can produce 200 pairs of skis per month, Plant 2 can produce 100 pairs of skis at per month, and Plant 3 can produce 50 pairs. Alpine Sports can thus produce 350 pairs of skis per month if it devotes its resources exclusively to ski production. In that case, it produces no snowboards.

Now suppose the firm decides to produce 100 snowboards. That will require shifting one of its plants out of ski production. Which one will it choose to shift? The sensible thing for it to do is to choose the plant in which snowboards have the lowest opportunity cost—Plant 3. It has an advantage not because it can produce more snowboards than the other plants (all the plants in this example are capable of producing up to 100 snowboards per month) but because it is the least productive plant for making skis. Producing a snowboard in Plant 3 requires giving up just half a pair of skis.

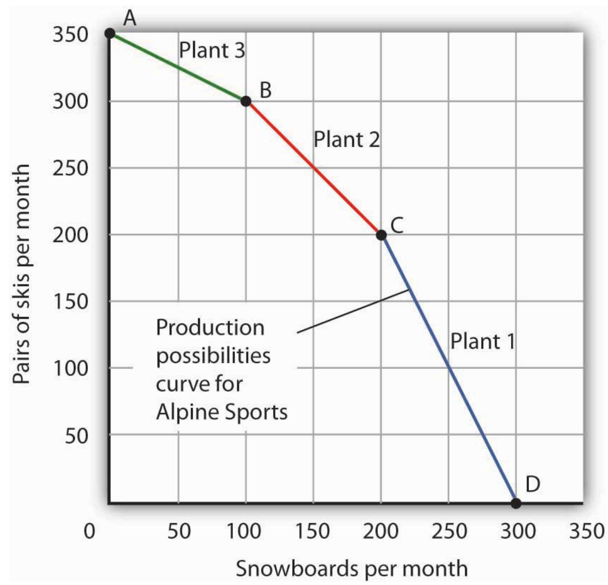
Economists say that an economy has a **comparative advantage** in producing a good or service if the opportunity cost of producing that good or service is lower for that economy than for any other. Plant 3 has a comparative advantage in snowboard production because it is the plant for which the opportunity cost of additional snowboards is lowest. To put this in terms of the production possibilities curve, Plant 3 has a comparative advantage in snowboard production (the good on the horizontal axis) because its production possibilities curve is the flattest of the three curves.

comparative advantage

In producing a good or service, the situation that occurs if the opportunity cost of producing that good or service is lower for that economy than for any other.

FIGURE 2.5 The Combined Production Possibilities Curve for Alpine Sports

The curve shown combines the production possibilities curves for each plant. At point A, Alpine Sports produces 350 pairs of skis per month and no snowboards. If the firm wishes to increase snowboard production, it will first use Plant 3, which has a comparative advantage in snowboards.



Plant 3's comparative advantage in snowboard production makes a crucial point about the nature of comparative advantage. It need not imply that a particular plant is especially good at an activity. In our example, all three plants are equally good at snowboard production. Plant 3, though, is the least efficient of the three in ski production. Alpine thus gives up fewer skis when it produces snowboards in Plant 3. Comparative advantage thus can stem from a lack of efficiency in the production of an alternative good rather than a special proficiency in the production of the first good.

The combined production possibilities curve for the firm's three plants is shown in Figure 2.5. We begin at point A, with all three plants producing only skis. Production totals 350 pairs of skis per month and zero snowboards. If the firm were to produce 100 snowboards at Plant 3, ski production would fall by 50 pairs per month (recall that the opportunity cost per snowboard at Plant 3 is half a pair of skis). That would bring ski production to 300 pairs, at point B. If Alpine Sports were to produce still more snowboards in a single month, it would shift production to Plant 2, the facility with the next-lowest opportunity cost. Producing 100 snowboards at Plant 2 would leave Alpine Sports producing 200 snowboards and 200 pairs of skis per month, at point C. If the firm were to switch entirely to snowboard production, Plant 1 would be the last to switch because the cost of each snowboard there is 2 pairs of skis. With all three plants producing only snowboards, the firm is at point D on the combined production possibilities curve, producing 300 snowboards per month and no skis.

Notice that this production possibilities curve, which is made up of linear segments from each assembly plant, has a bowed-out shape; the absolute value of its slope increases as Alpine Sports produces more and more snowboards. This is a result of transferring resources from the production of one good to another according to comparative advantage. We shall examine the significance of the bowed-out shape of the curve in the next section.

2.3 The Law of Increasing Opportunity Cost

law of increasing opportunity cost

As an economy moves along its production possibilities curve in the direction of producing more of a particular good, the opportunity cost of additional units of that good will increase.

We see in Figure 2.5 that, beginning at point A and producing only skis, Alpine Sports experiences higher and higher opportunity costs as it produces more snowboards. The fact that the opportunity cost of additional snowboards increases as the firm produces more of them is a reflection of an important economic law. The **law of increasing opportunity cost** holds that as an economy moves along its production possibilities curve in the direction of producing more of a particular good, the opportunity cost of additional units of that good will increase.

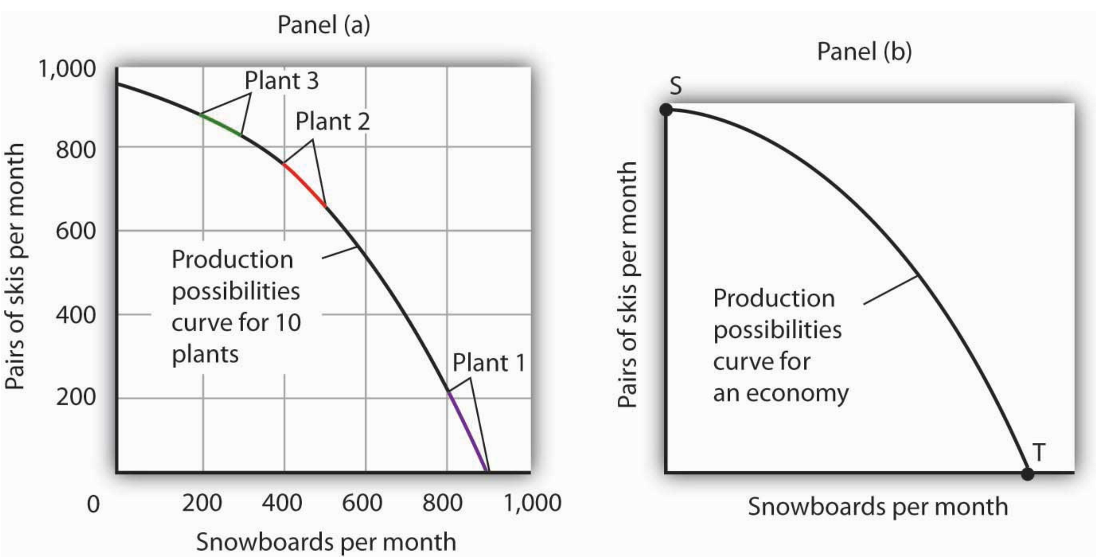
We have seen the law of increasing opportunity cost at work traveling from point A toward point D on the production possibilities curve in Figure 2.5. The opportunity cost of each of the first 100 snowboards equals half a pair of skis; each of the next 100 snowboards has an opportunity cost of 1 pair of skis, and each of the last 100 snowboards has an opportunity cost of 2 pairs of skis. The law also applies as the firm shifts from snowboards to skis. Suppose it begins at point D, producing 300 snowboards per month and no skis. It can shift to ski production at a relatively low cost at first. The opportunity cost of the first 200 pairs of skis is just 100 snowboards at Plant 1, a movement from point D to point C, or 0.5 snowboards per pair of skis. We would say that Plant 1 has a comparative advantage in ski production. The next 100 pairs of skis would be produced at Plant 2, where snowboard production would fall by 100 snowboards per month. The opportunity cost of skis at Plant 2 is 1 snowboard per pair of skis. Plant 3 would be the last plant converted to ski production. There, 50 pairs of skis could be produced per month at a cost of 100 snowboards, or an opportunity cost of 2 snowboards per pair of skis.

The bowed-out production possibilities curve for Alpine Sports illustrates the law of increasing opportunity cost. Scarcity implies that a production possibilities curve is downward sloping; the law of increasing opportunity cost implies that it will be bowed out, or concave, in shape.

The bowed-out curve of Figure 2.5 becomes smoother as we include more production facilities. Suppose Alpine Sports expands to 10 plants, each with a linear production possibilities curve. Panel (a) of Figure 2.6 shows the combined curve for the expanded firm, constructed as we did in Figure 2.5. This production possibilities curve includes 10 linear segments and is almost a smooth curve. As we include more and more production units, the curve will become smoother and smoother. In an actual economy, with a tremendous number of firms and workers, it is easy to see that the production possibilities curve will be smooth. We will generally draw production possibilities curves for the economy as smooth, bowed-out curves, like the one in Panel (b). This production possibilities curve shows an economy that produces only skis and snowboards. Notice the curve still has a bowed-out shape; it still has a negative slope. Notice also that this curve has no numbers. Economists often use models such as the production possibilities model with graphs that show the general shapes of curves but that do not include specific numbers.

FIGURE 2.6 Production Possibilities for the Economy

As we combine the production possibilities curves for more and more units, the curve becomes smoother. It retains its negative slope and bowed-out shape. In Panel (a) we have a combined production possibilities curve for Alpine Sports, assuming that it now has 10 plants producing skis and snowboards. Even though each of the plants has a linear curve, combining them according to comparative advantage, as we did with 3 plants in Figure 2.5, produces what appears to be a smooth, nonlinear curve, even though it is made up of linear segments. In drawing production possibilities curves for the economy, we shall generally assume they are smooth and "bowed out," as in Panel (b). This curve depicts an entire economy that produces only skis and snowboards.



2.4 Movements Along the Production Possibilities Curve

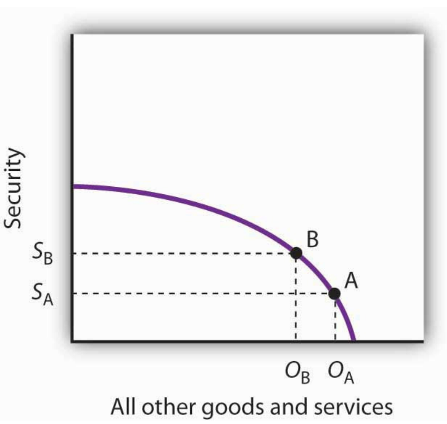
We can use the production possibilities model to examine choices in the production of goods and services. In applying the model, we assume that the economy can produce two goods, and we assume that technology and the factors of production available to the economy remain unchanged. In this section, we shall assume that the economy operates on its production possibilities curve so that an increase in the production of one good in the model implies a reduction in the production of the other.

We shall consider two goods and services: national security and a category we shall call "all other goods and services." This second category includes the entire range of goods and services the economy can produce, aside from national defense and security. Clearly, the transfer of resources to the effort to enhance national security reduces the quantity of other goods and services that can be produced. In the wake of the 9/11 attacks in 2001, nations throughout the world increased their spending for national security. This spending took a variety of forms. One, of course, was increased defense spending. Local and state governments also increased spending in an effort to prevent terrorist attacks. Airports around the world hired additional agents to inspect luggage and passengers.

The increase in resources devoted to security meant fewer "other goods and services" could be produced. In terms of the production possibilities curve in Figure 2.7, the choice to produce more security and less of other goods and services means a movement from A to B. Of course, an economy cannot really *produce* security; it can only attempt to provide it. The attempt to provide it requires resources; it is in that sense that we shall speak of the economy as "producing" security.

FIGURE 2.7 **Spending More for Security**

Here, an economy that can produce two categories of goods, security and "all other goods and services," begins at point A on its production possibilities curve. The economy produces S_A units of security and O_A units of all other goods and services per period. A movement from A to B requires shifting resources out of the production of all other goods and services and into spending on security. The increase in spending on security, to S_A units of security per period, has an opportunity cost of reduced production of all other goods and services. Production of all other goods and services falls by O_A - O_B units per period.



At point A, the economy was producing S_A units of security on the vertical axis—defense services and various forms of police protection—and O_A units of other goods and services on the horizontal axis. The decision to devote more resources to security and less to other goods and services represents the choice we discussed in the chapter introduction. In this case we have categories of goods rather than specific goods. Thus, the economy chose to increase spending on security in the effort to defeat terrorism. Since we have assumed that the economy has a fixed quantity of available resources, the increased use of resources for security and national defense necessarily reduces the number of resources available for the production of other goods and services.

The law of increasing opportunity cost tells us that, as the economy moves along the production possibilities curve in the direction of more of one good, its opportunity cost will increase. We may conclude that, as the economy moved along this curve in the direction of greater production of security, the opportunity cost of the additional security began to increase. That is because the resources transferred from the production of other goods and services to the production of security had a greater and greater comparative advantage in producing things other than security.

The production possibilities model does not tell us where on the curve a particular economy will operate. Instead, it lays out the possibilities facing the economy. Many countries, for example, chose to move along their respective production possibilities curves to produce more security and national defense and less of all other goods in the wake of 9/11. We will see in the chapter on demand and supply how choices about what to produce are made in the marketplace.

2.5 Producing on Versus Producing Inside the Production Possibilities Curve

An economy that is operating inside its production possibilities curve could, by moving onto it, produce more of all the goods and services that people value, such as food, housing, education, medical care, and music. Increasing the availability of these goods would improve the standard of living. Economists conclude that it is better to be on the production possibilities curve than inside it.

Two things could leave an economy operating at a point inside its production possibilities curve. First, the economy might fail to use fully the resources available to it. Second, it might not allocate resources on the basis of comparative advantage. In either case, production within the production possibilities curve implies the economy could improve its performance.

Idle Factors of Production

Suppose an economy fails to put all its factors of production to work. Some workers are without jobs, some buildings are without occupants, some fields are without crops. Because an economy's production possibilities curve assumes the full use of the factors of production available to it, the failure to use some factors results in a level of production that lies inside the production possibilities curve.

If all the factors of production that are available for use under current market conditions are being utilized, the economy has achieved **full employment**. An economy cannot operate on its production possibilities curve unless it has full employment.

full employment

Situation in which all the factors of production that are available for use under current market conditions are being utilized.

FIGURE 2.8 Idle Factors and Production

The production possibilities curve shown suggests an economy that can produce two goods, food and clothing. As a result of a failure to achieve full employment, the economy operates at a point such as B, producing F_B units of food and C_B units of clothing per period. Putting its factors of production to work allows a move to the production possibilities curve, to a point such as A. The production of both goods rises.

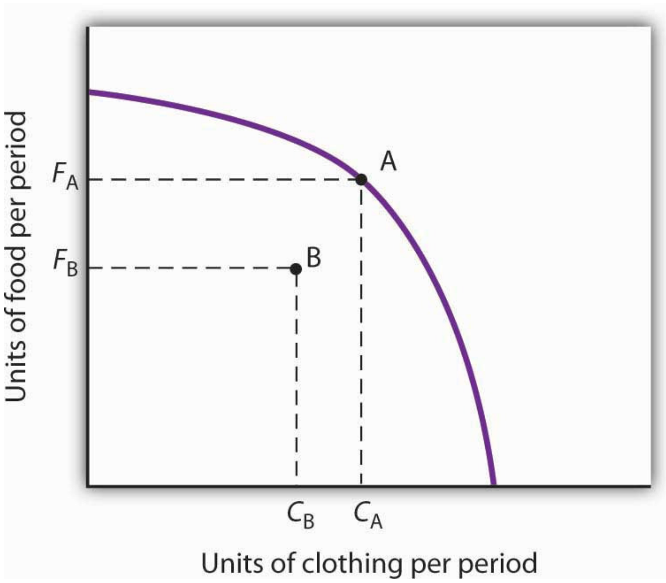


Figure 2.8 shows an economy that can produce food and clothing. If it chooses to produce at point A, for example, it can produce F_A units of food and C_A units of clothing. Now suppose that a large fraction of the economy's workers lose their jobs, so the economy no longer makes full use of one factor of production: labor. In this example, production moves to point B, where the economy produces less food (F_B) and less clothing (C_B) than at point A. We often think of the loss of jobs in terms of the workers; they have lost a chance to work and to earn income. But the production possibilities model points to another loss: goods and services the economy could have produced that are not being produced.

Inefficient Production

Now suppose Alpine Sports is fully employing its factors of production. Could it still operate inside its production possibilities curve? Could an economy that is using all its factors of production still produce less than it could? The answer is "Yes," and the key lies in comparative advantage. An economy achieves a point on its production possibilities curve only if it allocates its factors of production on the basis of comparative advantage. If it fails to do that, it will operate inside the curve.

Suppose that, as before, Alpine Sports has been producing only skis. With all three of its plants producing skis, it can produce 350 pairs of skis per month (and no snowboards). The firm then starts producing snowboards. This time, however, imagine that Alpine Sports switches plants from skis to snowboards in numerical order: Plant 1 first, Plant 2 second, and then Plant 3. Figure 2.9 illustrates the result. Instead of the bowed-out production possibilities curve ABCD, we get a bowed-in curve, AB′C′D. Suppose that Alpine Sports is producing 100 snowboards and 150 pairs of skis at point B′. Had the firm based its production choices on comparative advantage, it would have switched Plant 3 to snowboards and then Plant 2, so it could have operated at a point such as C. It would be producing more snowboards and more pairs of skis—and using the same quantities of factors of production it was using at B′. Had the firm based its production choices on comparative advantage, it would have switched Plant 3 to snowboards and then Plant 2, so it would have operated at point C. It would be producing more snowboards and more pairs of skis—and using the same quantities of factors of production it was using at B′. When an economy is operating on its production possibilities curve, we say that it is engaging in **efficient production**. If it is using the same quantities of factors of production but is operating inside its production possibilities curve, it is engaging in **inefficient production**. Inefficient production implies that the economy could be producing more goods without using any additional labor, capital, or natural resources.

efficient production

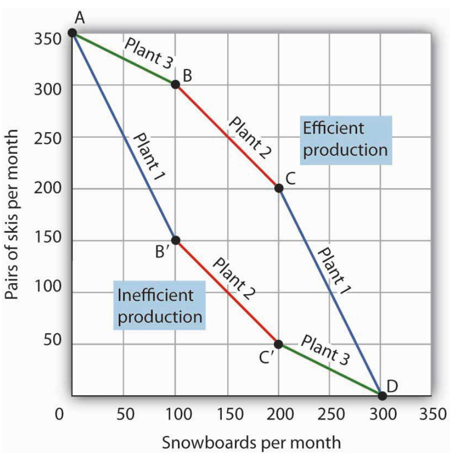
When an economy is operating on its production possibilities curve.

inefficient production

Situation in which the economy is using the same quantities of factors of production but is operating inside its production possibilities curve.

FIGURE 2.9 Efficient Versus Inefficient Production

When factors of production are allocated on a basis other than comparative advantage, the result is inefficient production. Suppose Alpine Sports operates the three plants we examined in Figure 2.4. Suppose further that all three plants are devoted exclusively to ski production; the firm operates at A. Now suppose that, to increase snowboard production, it transfers plants in numerical order: Plant 1 first, then Plant 2, and finally Plant 3. The result is the bowed-in curve AB′C′D. Production on the production possibilities curve ABCD requires that factors of production be transferred according to comparative advantage.



Points on the production possibilities curve thus satisfy two conditions: the economy is making full use of its factors of production, and it is making efficient use of its factors of production. If there are idle or inefficiently allocated factors of production, the economy will operate inside the production possibilities curve. Thus, the production possibilities curve not only shows what can be produced; it provides insight into how goods and services should be produced. It suggests that to obtain efficiency in production, factors of production should be allocated on the basis of comparative advantage. Further, the economy must make full use of its factors of production if it is to produce the goods and services it is capable of producing.

Specialization

The production possibilities model suggests that specialization will occur. **Specialization** implies that an economy is producing the goods and services in which it has a comparative advantage. If Alpine Sports selects point C in Figure 2.9, for example, it will assign Plant 1 exclusively to ski production and Plants 2 and 3 exclusively to snowboard production.

Such specialization is typical in an economic system. Workers, for example, specialize in particular fields in which they have a comparative advantage. People work and use the income they earn to buy—perhaps import—goods and services from people who have a comparative advantage in doing other things. The result is a far greater quantity of goods and services than would be available without this specialization.

Think about what life would be like without specialization. Imagine that you are suddenly completely cut off from the rest of the economy. You must produce everything you consume; you obtain nothing from anyone else. Would you be able to consume what you consume now? Clearly not. It is hard to imagine that most of us could even survive in such a setting. The gains we achieve through specialization are enormous.

Nations specialize as well. Much of the land in the United States has a comparative advantage in agricultural production and is devoted to that activity. Hong Kong, with its huge population and tiny endowment of land, allocates virtually none of its land to agricultural use; that option would be too costly. Its land is devoted largely to nonagricultural use.

specialization

Situation in which an economy is producing the goods and services in which it has a comparative advantage.

KEY TAKEAWAYS

- A production possibilities curve shows the combinations of two goods an economy is capable of producing.
- The downward slope of the production possibilities curve is an implication of scarcity.
- The bowed-out shape of the production possibilities curve results from allocating resources based on comparative advantage. Such an allocation implies that the law of increasing opportunity cost will hold.
- An economy that fails to make full and efficient use of its factors of production will operate inside its production possibilities curve.
- Specialization means that an economy is producing the goods and services in which it has a comparative advantage.

TRY IT!

Suppose a manufacturing firm is equipped to produce radios or calculators. It has two plants, Plant R and Plant S, at which it can produce these goods. Given the labor and the capital available at both plants, it can produce the combinations of the two goods at the two plants shown.

Output per day, Plant R

Combination	Calculators	Radios
A	100	0
B	50	25
C	0	50

Output per day, Plant S

Combination	Calculators	Radios
D	50	0
E	25	50
F	0	100

Put calculators on the vertical axis and radios on the horizontal axis. Draw the production possibilities curve for Plant R. On a separate graph, draw the production possibilities curve for Plant S. Which plant has a comparative advantage in calculators? In radios? Now draw the combined curves for the two plants. Suppose the firm decides to produce 100 radios. Where will it produce them? How many calculators will it be able to produce? Where will it produce the calculators?

Case in Point: The Cost of the Great Depression

The U.S. economy looked very healthy in the beginning of 1929. It had enjoyed seven years of dramatic growth and unprecedented prosperity. Its resources were fully employed; it was operating quite close to its production possibilities curve.

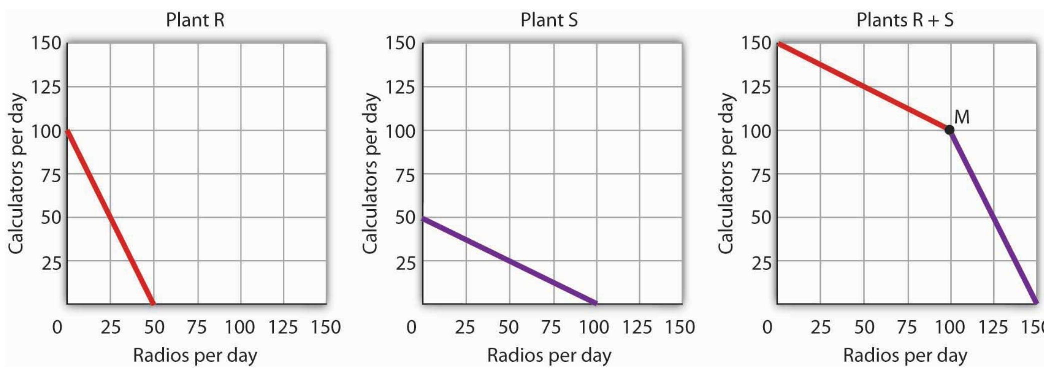
In the summer of 1929, however, things started going wrong. Production and employment fell. They continued to fall for several years. By 1933, more than 25% of the nation's workers had lost their jobs. Production had plummeted by almost 30%. The economy had moved well within its production possibilities curve.

Output began to grow after 1933, but the economy continued to have vast numbers of idle workers, idle factories, and idle farms. These resources were not put back to work fully until 1942, after the U.S. entry into World War II demanded mobilization of the economy's factors of production.

Between 1929 and 1942, the economy produced 25% fewer goods and services than it would have if its resources had been fully employed. That was a loss, measured in today's dollars, of well over $3 trillion. In material terms, the forgone output represented a greater cost than the United States would ultimately spend in World War II. The Great Depression was a costly experience indeed.

ANSWER TO TRY IT! PROBLEM

The production possibilities curves for the two plants are shown, along with the combined curve for both plants. Plant R has a comparative advantage in producing calculators. Plant S has a comparative advantage in producing radios, so, if the firm goes from producing 150 calculators and no radios to producing 100 radios, it will produce them at Plant S. In the production possibilities curve for both plants, the firm would be at M, producing 100 calculators at Plant R.



3. APPLICATIONS OF THE PRODUCTION POSSIBILITIES MODEL

LEARNING OBJECTIVES

1. **Understand the argument for unrestricted international trade in terms of economic specialization and comparative advantage.**
2. **Define economic growth in terms of the production possibilities model and discuss factors that make such growth possible.**
3. **Explain the classification of economic systems, the role of government in different economic systems, and the strengths and weaknesses of different systems.**

The production possibilities curve gives us a model of an economy. The model provides powerful insights about the real world, insights that help us to answer some important questions: How does trade between two countries affect the quantities of goods available to people? What determines the rate at which production will increase over time? What is the role of economic freedom in the economy? In this section we explore applications of the model to questions of international trade, economic growth, and the choice of an economic system.

3.1 Comparative Advantage and International Trade

One of the most important implications of the concepts of comparative advantage and the production possibilities curve relates to international trade. We can think of different nations as being equivalent to Christie Ryder's plants. Each will have a comparative advantage in certain activities, and efficient world production requires that each nation specialize in those activities in which it has a comparative advantage. A failure to allocate resources in this way means that world production falls inside the production possibilities curve; more of each good could be produced by relying on comparative advantage.

If nations specialize, then they must rely on each other. They will sell the goods in which they specialize and purchase other goods from other nations. Suppose, for example, that the world consists of two continents that can each produce two goods: South America and Europe can produce food and computers. Suppose they can produce the two goods according to the tables in Panels (a) and (b) of Figure 2.12. We have simplified this example by assuming that each continent has a linear production possibilities curve; the curves are plotted below the tables in Panels (a) and (b). Each continent has a separate production possibilities curve; the two have been combined to illustrate a world production possibilities curve in Panel (c) of the exhibit.

FIGURE 2.12 Production Possibilities Curves and Trade

Suppose the world consists of two continents: South America and Europe. They can each produce two goods: food and computers. In this example, we assume that each continent has a linear production possibilities curve, as shown in Panels (a) and (b). South America has a comparative advantage in food production and Europe has a comparative advantage in computer production. With free trade, the world can operate on the bowed-out curve GHI, shown in Panel (c). If the continents refuse to trade, the world will operate inside its production possibilities curve. If, for example, each continent were to produce at the midpoint of its production possibilities curve, the world would produce 300 computers and 300 units of food per period at point Q. If each continent were to specialize in the good in which it has a comparative advantage, world production could move to a point such as H, with more of both goods produced.

Panel (a)
South America's Production Possibilities

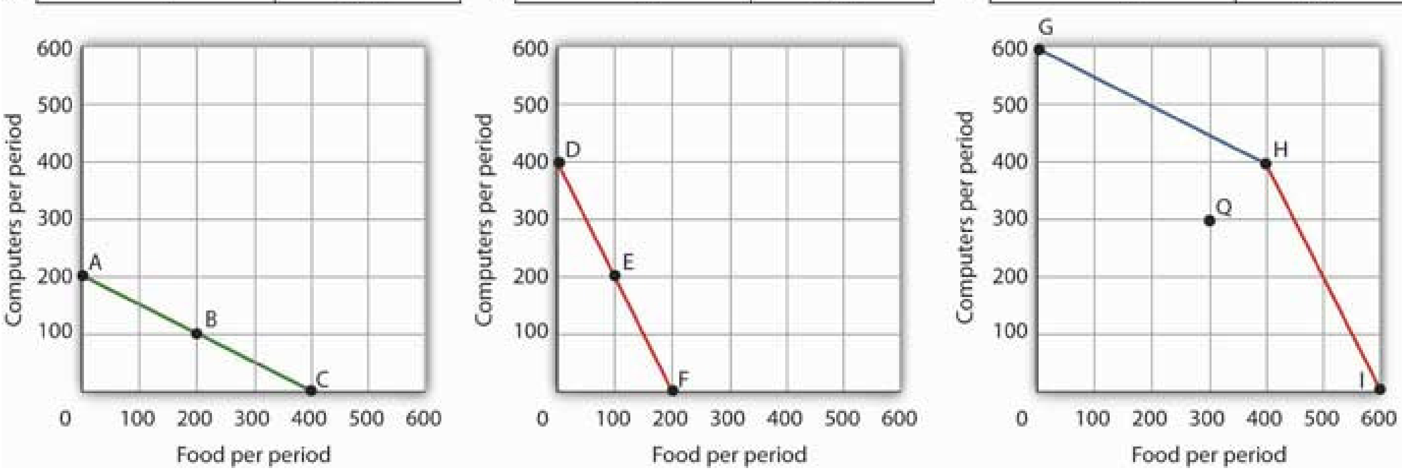
	Computers per period	Food per period
A	200	0
B	100	200
C	0	400

Panel (b)
Europe's Production Possibilities

	Computers per period	Food per period
D	400	0
E	200	100
F	0	200

Panel (c)
World Production Possibilities

	Computers per period	Food per period
G	600	0
H	400	400
I	0	600



The world production possibilities curve assumes that resources are allocated between computer and food production based on comparative advantage. Notice that, even with only two economies and the assumption of linear production possibilities curves for each, the combined curve still has a bowed-out shape. At point H, for example, South America specializes in food, while Europe produces only computers. World production equals 400 units of each good. In this situation, we would expect South America to export food to Europe while Europe exports computers to South America.

But suppose the regions refuse to trade; each insists on producing its own food and computers. Suppose further that each chooses to produce at the midpoint of its own production possibilities curve. South America produces 100 units of computers and 200 units of food per period, while Europe produces 200 units of computers and 100 units of food per period. World production thus totals 300 units of each good per period; the world operates at point Q in Figure 2.12. If the two continents were willing to move from isolation to trade, the world could achieve an increase in the production of both goods. Producing at point H requires no more resources, no more effort than production at Q. It does, however, require that the world's resources be allocated on the basis of comparative advantage.

The implications of our model for trade are powerful indeed. First, we see that trade allows the production of more of all goods and services. Restrictions on trade thus reduce production of goods and services. Second, we see a lesson often missed in discussions of trade: a nation's trade policy has nothing to do with its level of employment of its factors of production. In our example, when South America and Europe do not engage in trade and produce at the midpoints of each of their respective production possibilities curves, *they each have full employment*. With trade, the two nations still operate on their respective production possibilities curves: *they each have full employment*. Trade certainly redistributes employment in the two continents. In South America, employment shifts from computer production to food production. In Europe, it shifts from food production to computer production. Once the shift is made, though, there is no effect on employment in either continent.

Of course, this idealized example would have all of South America's computer experts becoming farmers while all of Europe's farmers become computer geeks! That is a bit much to swallow, but it is merely the result of assuming linear production possibilities curves and complete specialization. In the real world, production possibilities curves are concave, and the reallocation of resources required by trade is not nearly as dramatic. Still, free trade can require shifts in resources from one activity to another. These shifts produce enormous benefits, but they do not come without costs.

Nearly all economists agree that largely unrestricted trade between countries is desirable; restrictions on trade generally force the world to operate inside its production possibilities curve. In some cases restrictions on trade could be desirable, but in the main, free trade promotes greater production of goods and services for the world's people. The role of international trade is explored in greater detail in subsequent chapters of this book.

3.2 Economic Growth

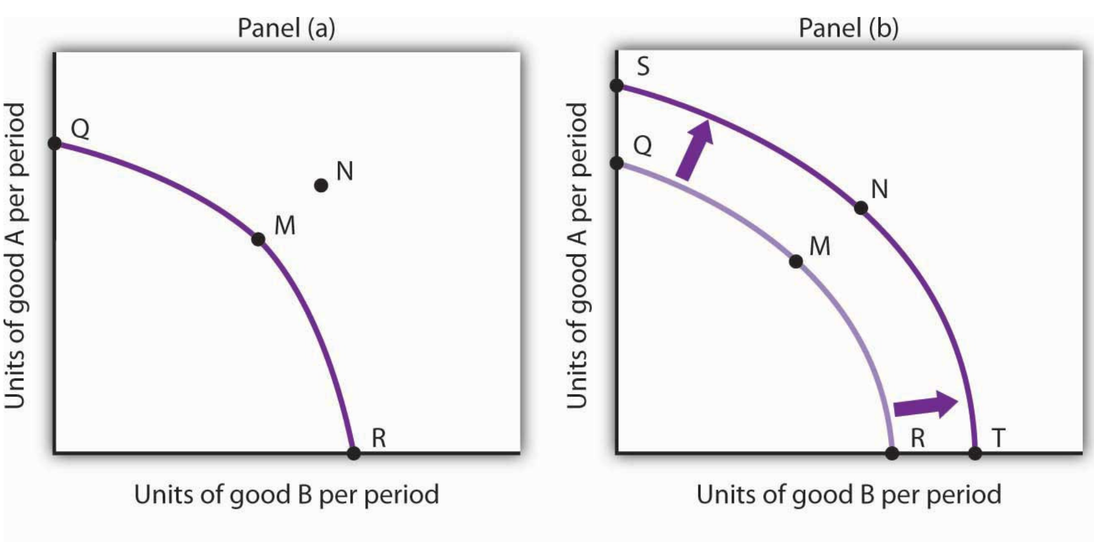
economic growth

The process through which an economy achieves an outward shift in its production possibilities curve.

An increase in the physical quantity or in the quality of factors of production available to an economy or a technological gain will allow the economy to produce more goods and services; it will shift the economy's production possibilities curve outward. The process through which an economy achieves an outward shift in its production possibilities curve is called **economic growth**. An outward shift in a production possibilities curve is illustrated in Figure 2.13. In Panel (a), a point such as N is not attainable; it lies outside the production possibilities curve. Growth shifts the curve outward, as in Panel (b), making previously unattainable levels of production possible.

FIGURE 2.13 Economic Growth and the Production Possibilities Curve

An economy capable of producing two goods, A and B, is initially operating at point M on production possibilities curve OMR in Panel (a). Given this production possibilities curve, the economy could not produce a combination such as shown by point N, which lies outside the curve. An increase in the factors of production available to the economy would shift the curve outward to SNT, allowing the choice of a point such as N, at which more of both goods will be produced.



The Sources of Economic Growth

Economic growth implies an outward shift in an economy's production possibilities curve. Recall that when we draw such a curve, we assume that the quantity and quality of the economy's factors of production and its technology are unchanged. Changing these will shift the curve. Anything that increases the quantity or quality of the factors of production available to the economy or that improves the technology available to the economy contributes to economic growth.

Consider, for example, the dramatic gains in human capital that have occurred in the United States since the beginning of the past century. In 1900, about 3.5% of U.S. workers had completed a high school education. By 2006, that percentage rose almost to 92. Fewer than 1% of the workers in 1900 had graduated from college; as late as 1940 only 3.5% had graduated from college. By 2006, nearly 32% had graduated from college. In addition to being better educated, today's workers have received more and better training on the job. They bring far more economically useful knowledge and skills to their work than did workers a century ago.

Moreover, the technological changes that have occurred within the past 100 years have greatly reduced the time and effort required to produce most goods and services. Automated production has become commonplace. Innovations in transportation (automobiles, trucks, and airplanes) have made the movement of goods and people cheaper and faster. A dizzying array of new materials is available for manufacturing. And the development of modern information technology—including computers, software, and communications equipment—that seemed to proceed at breathtaking pace especially during the final years of the last century and continuing to the present has transformed the way we live and work.

Look again at the technological changes of the last few years described in the Case in Point on advances in technology. Those examples of technological progress through applications of computer technology—from new ways of mapping oil deposits to new methods of milking cows—helped propel the United States and other economies to dramatic gains in the ability to produce goods and services. They have helped shift the countries' production possibilities curve outward. They have helped fuel economic growth.

Table 2.1 summarizes the factors that have contributed to U.S. economic growth in the past half century. When looking at the period of 1948–2002 as a whole we see that about 60% of economic growth stems from increases in the quantities of capital and labor and 40% from increases in the qualities of the factors of production and improvements in technology. In the most recent period, 1995–2002, however, these percentages are essentially reversed, with a little less than 30% explained by increases in quantities of the factors of production and a whopping 70% explained by improvements in factor quality and technology.

TABLE 2.1 Sources of U.S. Economic Growth, 1948–2002

Total output during the period shown increased sixfold. The chart shows the percentage of this increase accounted for by increases in the quantity of labor and of capital and by increases in the quality of labor and of capital and improvements in technology. In the 1995–2002 period, the incorporation of information technology led to improvements in the quality of capital and technology that greatly contributed to growth.

Year	Percentage contribution to growth	Period growth rate
Years 1948–2002		**3.46%**
Increase in quantity of labor	21%	
Increase in quantity of capital	41%	
Increase in quality of labor	10%	
Increase in quality of capital	20%	
Improved technology	25%	
Years 1948–1973		**3.99%**
Increase in quantity of labor	15%	
Increase in quantity of capital	44%	
Increase in quality of labor	11%	
Increase in quality of capital	5%	
Improved technology	25%	
Years 1973–1989		**2.97%**
Increase in quantity of labor	31%	
Increase in quantity of capital	39%	
Increase in quality of labor	7%	
Increase in quality of capital	12%	
Improved technology	10%	
Years 1989–1995		**2.43%**
Increase in quantity of labor	26%	
Increase in quantity of capital	33%	
Increase in quality of labor	15%	
Increase in quality of capital	17%	
Improved technology	11%	
Years 1995–2002		**3.59%**
Increase in quantity of labor	19%	
Increase in quantity of capital	8%	
Increase in quality of labor	5%	
Increase in quality of capital	47%	
Improved technology	20%	

Source: Based on Dale W. Jorgenson, "Accounting for Growth in the Information Age," Handbook of Economic Growth, Phillipe Aghion and Steven Durlauf, eds. Amsterdam: North Holland, 2005.

Another way of looking at these data for the most recent period is to notice that the increase in the rate of economic growth between the 1989 to 1995 period and the 1995 to 2002 period of more than one percentage point per year is largely explained by better-quality capital and better technology. The study by economist Dale Jorgenson on which the data shown in Table 2.1 are derived notes that these two main contributors to higher economic growth can be largely attributed to the development of information technology and its incorporation in the workplace.

Waiting for Growth

One key to growth is, in effect, the willingness to wait, to postpone current consumption in order to enhance future productive capability. When Stone Age people fashioned the first tools, they were spending time building capital rather than engaging in consumption. They delayed current consumption to enhance their future consumption; the tools they made would make them more productive in the future.

Resources society could have used to produce consumer goods are being used to produce new capital goods and new knowledge for production instead—all to enhance future production. An even more important source of growth in many nations has been increased human capital. Increases in human capital often require the postponement of consumption. If you are a college student, you are engaged in precisely this effort. You are devoting time to study that could have been spent working, earning income, and thus engaging in a higher level of consumption. If you are like most students, you are making this choice to postpone consumption because you expect it will allow you to earn more income, and thus enjoy greater consumption, in the future.

Think of an economy as being able to produce two goods, capital and consumer goods (those destined for immediate use by consumers). By focusing on the production of consumer goods, the people in the economy will be able to enjoy a higher standard of living today. If they reduce their consumption—and their standard of living—today to enhance their ability to produce goods and services in the future, they will be able to shift their production possibilities curve outward. That may allow them to produce even more consumer goods. A decision for greater growth typically involves the sacrifice of present consumption.

3.3 Arenas for Choice: A Comparison of Economic Systems

Under what circumstances will a nation achieve efficiency in the use of its factors of production? The discussion above suggested that Christie Ryder would have an incentive to allocate her plants efficiently because by doing so she could achieve greater output of skis and snowboards than would be possible from inefficient production. But why would she want to produce more of these two goods—or of any goods? Why would decision makers throughout the economy want to achieve such efficiency?

Economists assume that privately owned firms seek to maximize their profits. The drive to maximize profits will lead firms such as Alpine Sports to allocate resources efficiently to gain as much production as possible from their factors of production. But whether firms will seek to maximize profits depends on the nature of the economic system within which they operate.

Classifying Economic Systems

Each of the world's economies can be viewed as operating somewhere on a spectrum between market capitalism and command socialism. In a **market capitalist economy**, resources are generally owned by private individuals who have the power to make decisions about their use. A market capitalist system is often referred to as a free enterprise economic system. In a **command socialist economy**, the government is the primary owner of capital and natural resources and has broad power to allocate the use of factors of production. Between these two categories lie **mixed economies** that combine elements of market capitalist and of command socialist economic systems.

No economy represents a pure case of either market capitalism or command socialism. To determine where an economy lies between these two types of systems, we evaluate the extent of government ownership of capital and natural resources and the degree to which government is involved in decisions about the use of factors of production.

The diagram below suggests the spectrum of economic systems. Market capitalist economies lie toward the left end of this spectrum; command socialist economies appear toward the right. Mixed economies lie in between. The market capitalist end of the spectrum includes countries such as the United States, the United Kingdom, and Chile. Hong Kong, though now part of China, has a long history as a market capitalist economy and is generally regarded as operating at the market capitalist end of the spectrum. Countries at the command socialist end of the spectrum include North Korea and Cuba.

market capitalist economy

Economy in which resources are generally owned by private individuals who have the power to make decisions about their use.

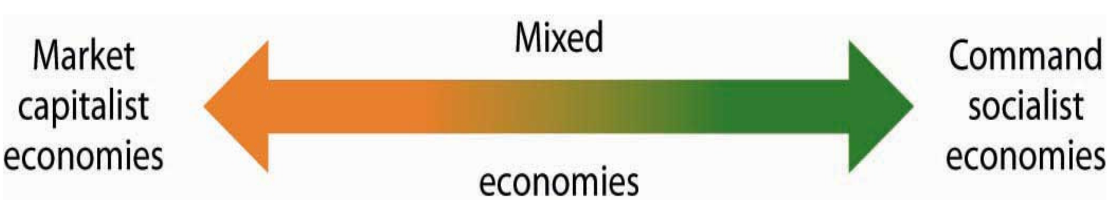
command socialist economy

Economy in which government is the primary owner of capital and natural resources and has broad power to allocate the use of factors of production.

mixed economies

Economy that combine elements of market capitalist and of command socialist economic systems.

FIGURE 2.14 Economic Systems



Some European economies, such as France, Germany, and Sweden, have a sufficiently high degree of regulation that we consider them as operating more toward the center of the spectrum. Russia and China, which long operated at the command socialist end of the spectrum, can now be considered mixed economies. Most economies in Latin America once operated toward the right end of the spectrum. While their governments did not exercise the extensive ownership of capital and natural resources that are one characteristic of command socialist systems, their governments did impose extensive regulations. Many of these nations are in the process of carrying out economic reforms that will move them further in the direction of market capitalism.

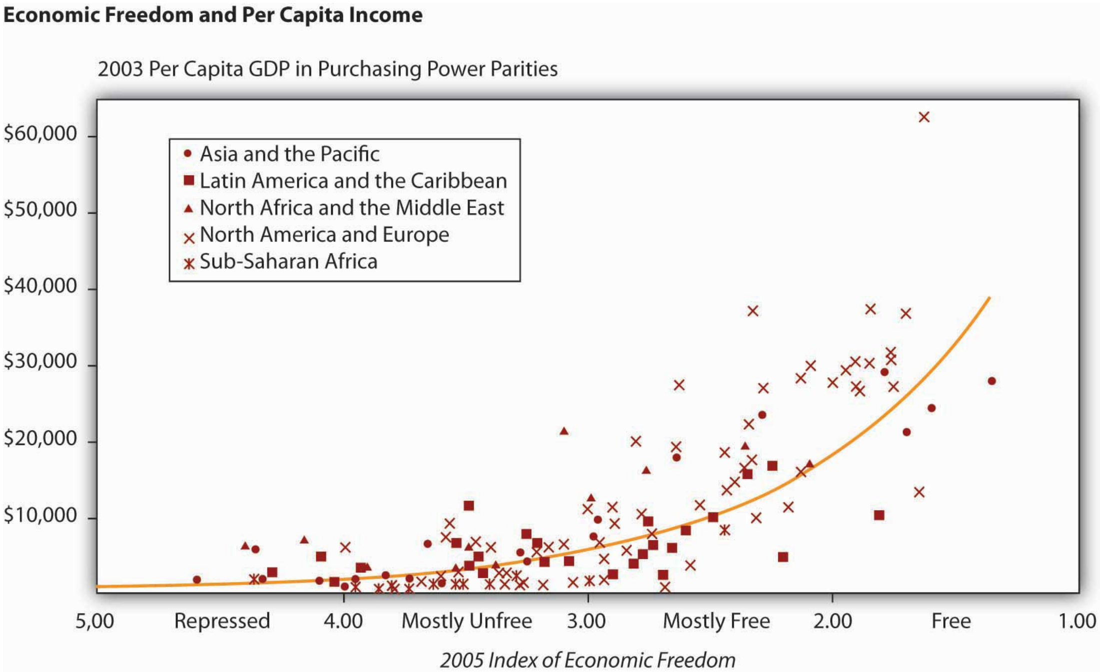
The global shift toward market capitalist economic systems that occurred in the 1980s and 1990s was in large part the result of three important features of such economies. First, the emphasis on individual ownership and decision-making power has generally yielded greater individual freedom than has been available under command socialist or some more heavily regulated mixed economic systems that lie toward the command socialist end of the spectrum. People seeking political, religious, and economic freedom have thus gravitated toward market capitalism. Second, market economies are more likely than other systems to allocate resources on the basis of comparative advantage. They thus tend to generate higher levels of production and income than do other economic systems. Third, market capitalist-type systems appear to be the most conducive to entrepreneurial activity.

Suppose Christie Ryder had the same three plants we considered earlier in this chapter but was operating in a mixed economic system with extensive government regulation. In such a system, she might be prohibited from transferring resources from one use to another to achieve the gains possible from comparative advantage. If she were operating under a command socialist system, she would not be the owner of the plants and thus would be unlikely to profit from their efficient use. If that were the case, there is no reason to believe she would make any effort to assure the efficient use of the three plants. Generally speaking, it is economies toward the market capitalist end of the spectrum that offer the greatest inducement to allocate resources on the basis of comparative advantage. They tend to be more productive and to deliver higher material standards of living than do economies that operate at or near the command socialist end of the spectrum.

Market capitalist economies rely on economic freedom. Indeed, one way we can assess the degree to which a country can be considered market capitalist is by the degree of economic freedom it permits. Several organizations have attempted to compare economic freedom in various countries. One of the most extensive comparisons is a joint annual effort by the Heritage Foundation and *The Wall Street Journal*. The 2008 rating was based on policies in effect in 162 nations early that year. The report ranks these nations on the basis of such things as the degree of regulation of firms, tax levels, and restrictions on international trade. Hong Kong ranked as the freest economy in the world. North Korea received the dubious distinction of being the least free.

FIGURE 2.15 Economic Freedom and Income

The horizontal axis shows the degree of economic freedom—"free," "mostly free," "mostly unfree," and "repressed"—according to the measures used by the Heritage Foundation and *The Wall Street Journal*. The graph shows the relationship between economic freedom and per capita income. Countries with higher degrees of economic freedom tended to have higher per capita incomes.



Source: World Bank, World Development Indicators Online, available by subscription at www.worldbank.org/data; Central Intelligence Agency, The World Factbook 2004, available at http://www.cia.gov/cia/publications/factbook/index.html for the following countries: Bahamas, Burma, Cuba, Cyprus, Equatorial Guinea, North Korea, Libya, Qatar, Suriname, Taiwan, Zimbabwe; Marc A. Miles, Edwin J. Feulner, and Mary Anastasia O'Grady, 2005 Index of Economic Freedom (Washington, D.C.: The Heritage Foundation and Dow Jones & Company, Inc., 2005), at www.heritage.org/index.

It seems reasonable to expect that the greater the degree of economic freedom a country permits, the greater the amount of income per person it will generate. This proposition is illustrated in Figure 2.15. The group of countries categorized as "free" generated the highest incomes in the Heritage Foundation/*Wall Street Journal* study; those rated as "repressed" had the lowest. The study also found that countries that over the last decade have done the most to improve their positions in the economic freedom rankings have also had the highest rates of growth. We must be wary of slipping into the fallacy of false cause by concluding from this evidence that economic freedom generates higher incomes. It could be that higher incomes lead nations to opt for greater economic freedom. But in this case, it seems reasonable to conclude that, in general, economic freedom does lead to higher incomes.

3.4 Government in a Market Economy

The production possibilities model provides a menu of choices among alternative combinations of goods and services. Given those choices, which combinations will be produced?

In a market economy, this question is answered in large part through the interaction of individual buyers and sellers. As we have already seen, government plays a role as well. It may seek to encourage greater consumption of some goods and discourage consumption of others. In the United States, for example, taxes imposed on cigarettes discourage smoking, while special treatment of property taxes and mortgage interest in the federal income tax encourages home ownership. Government may try to stop the production and consumption of some goods altogether, as many governments do with drugs such as heroin and cocaine. Government may supplement the private consumption of some goods by producing more of them itself, as many U.S. cities do with golf courses and tennis courts. In other cases, there may be no private market for a good or service at all. In the choice between security and defense versus all other goods and services outlined at the beginning of this chapter, government agencies are virtually the sole providers of security and national defense.

All nations also rely on government to provide defense, enforce laws, and redistribute income. Even market economies rely on government to regulate the activities of private firms, to protect the environment, to provide education, and to produce a wide range of other goods and services. Government's role may be limited in a market economy, but it remains fundamentally important.

KEY TAKEAWAYS

- The ideas of comparative advantage and specialization suggest that restrictions on international trade are likely to reduce production of goods and services.
- Economic growth is the result of increasing the quantity or quality of an economy's factors of production and of advances in technology.
- Policies to encourage growth generally involve postponing consumption to increase capital and human capital.
- Market capitalist economies have generally proved more productive than mixed or command socialist economies.
- Government plays a crucial role in any market economy.

TRY IT!

Draw a production possibilities curve for an economy that can produce two goods, CD players and jackets. You do not have numbers for this one—just draw a curve with the usual bowed-out shape. Put the quantity of CD players per period on the vertical axis and the quantity of jackets per period on the horizontal axis. Now mark a point A on the curve you have drawn; extend dotted lines from this point to the horizontal and vertical axes. Mark the initial quantities of the two goods as CD_A and J_A, respectively. Explain why, in the absence of economic growth, an increase in jacket production requires a reduction in the production of CD players. Now show how economic growth could lead to an increase in the production of both goods.

Case in Point: The European Union and the Production Possibilities Curve

Formed by the Maastricht Treaty of 1993, The European Union represents one of the boldest efforts of our time to exploit the theory of comparative advantage. The Treaty sought to eliminate all trade barriers between the European Union's members. It established a European Parliament and a European Central Bank. The Bank introduced the euro in 1999, a currency that replaced national currencies such as the German deutsche mark and the French franc. At first, the euro was used only for transactions between banks. 320 million people in 15 EU nations (Austria, Belgium, Cyprus, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Malta, the Netherlands, Portugal, Slovenia, and Spain) used the euro by 2008. While the dollar continues to be more widely used, the total value of euros in circulation exceeds that of dollars.

The movement toward European integration can be dated back more than half a century. In 1950, just five years after a war that had devastated much of the world, Robert Schuman, the French Minister of Foreign Affairs, proposed a union between France and Germany to cooperate in the production of iron and steel. In the context of the time, Schuman's proposal was a radical one. World War II had begun with Germany's attempt to seize control of Europe—and ultimately the world. Japan and Italy joined Germany in this effort. Germany had captured France; France had been liberated in 1944 by the Allied invasion in Normandy. The proposal for cooperation between two countries that had been the most bitter of enemies was a revolutionary one. Schuman's speech, delivered on May 9, 1950, is celebrated throughout Europe as "Europe Day."

In effect, the European Union has created an entity very much like the United States. Countries within the European Union retain their own languages and cultural differences, but they have ceded a remarkable degree of sovereignty to the Union. Members of the European Union can trade as freely with each other as can states within the United States. Just as the U.S. Constitution prohibits states from restricting trade with other states, the European Union has dismantled all forms of restrictions that countries within the Union used to impose on one another. Just as restrictions on specialization among Ms. Ryder's plants in Alpine Sports would have forced it to operate inside its production possibilities curve, restrictions that had existed among members of the European Union once put the members of the Union inside their collective production possibilities curve.

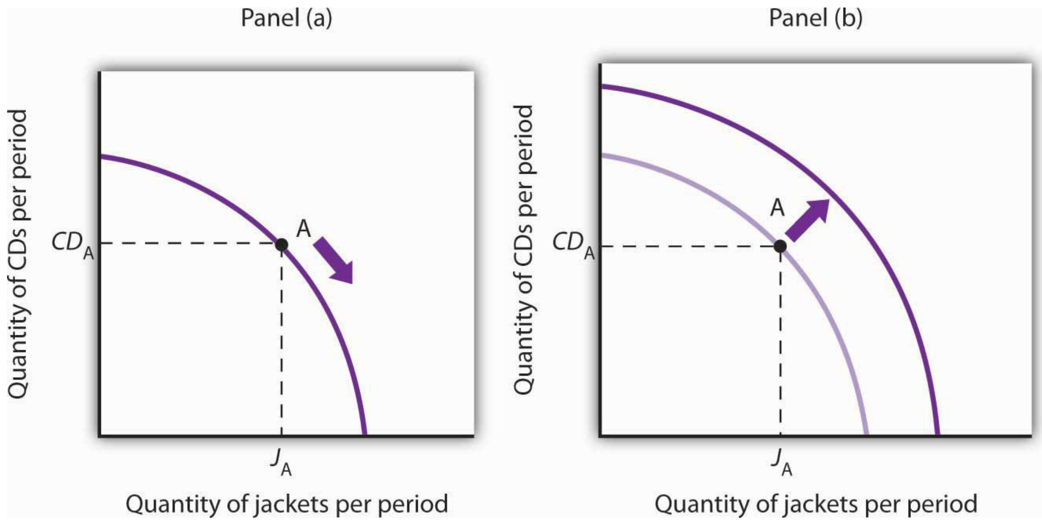
The experiment appears to have been a success. Trade among member nations has expanded sharply. A study by Carmen Diaz Mora, an economist at the University of Castilla-La Mancha in Spain, found that the bulk of the expanded trade within the Union was trade within industries and that it was driven by comparative advantage. In particular, she found that countries in the northern part of the Union, such as France and Germany, tended to specialize in relatively high-valued goods—office equipment and electrical goods—while countries in the southern part of the Union specialized in relatively low-valued goods such as food and textile products. In trade within the clothing industry, countries such as Italy tend to specialize in the production of higher-valued clothing, while lower-income countries such as Portugal specialize in the production of cheaper clothing. In sparkling wines, France specializes in the higher-quality end of the spectrum, while Spain specializes in the low-quality end. Similarly, Germany specializes in the production of higher-quality cars while Spain specializes in lower-quality vehicles. Similar exchanges occur across a wide range of goods and services.

Diaz Mora found that comparative advantage tended to correspond to income levels. Countries in the northern part of the European Union tend to have high per capita incomes and high levels of human capital and technology—these countries gained by specializing in the production of high-valued goods. Countries in the southern part of the Union also gained by specialization—in the production of low-valued goods. This specialization has increased the welfare of people throughout the Union.

Sources: Carmen Diaz Mora, "The Role of Comparative Advantage in Trade Within Industries: A Panel Data Approach for the European Union," Weltwirtschaftliches Archiv 138:2 (2002), 291–316.

ANSWER TO TRY IT! PROBLEM

Your first production possibilities curve should resemble the one in Panel (a). Starting at point A, an increase in jacket production requires a move down and to the right along the curve, as shown by the arrow, and thus a reduction in the production of CD players. Alternatively, if there is economic growth, it shifts the production possibilities curve outward, as in Panel (b). This shift allows an increase in production of both goods, as suggested by the arrow.



4. REVIEW AND PRACTICE

Summary

Economics deals with choices. In this chapter we have examined more carefully the range of choices in production that must be made in any economy. In particular, we looked at choices involving the allocation of an economy's factors of production: labor, capital, and natural resources.

In addition, in any economy, the level of technology plays a key role in determining how productive the factors of production will be. In a market economy, entrepreneurs organize factors of production and act to introduce technological change.

The production possibilities model is a device that assists us in thinking about many of the choices about resource allocation in an economy. The model assumes that the economy has factors of production that are fixed in both quantity and quality. When illustrated graphically, the production possibilities model typically limits our analysis to two goods. Given the economy's factors of production and technology, the economy can produce various combinations of the two goods. If it uses its factors of production efficiently and has full employment, it will be operating on the production possibilities curve.

Two characteristics of the production possibilities curve are particularly important. First, it is downward sloping. This reflects the scarcity of the factors of production available to the economy; producing more of one good requires giving up some of the other. Second, the curve is bowed out. Another way of saying this is to say that the curve gets steeper as we move from left to right; the absolute value of its slope is increasing. Producing each additional unit of the good on the horizontal axis requires a greater sacrifice of the good on the vertical axis than did the previous units produced. This fact, called the law of increasing opportunity cost, is the inevitable result of efficient choices in production—choices based on comparative advantage.

The production possibilities model has important implications for international trade. It suggests that free trade will allow countries to specialize in the production of goods and services in which they have a comparative advantage. This specialization increases the production of all goods and services.

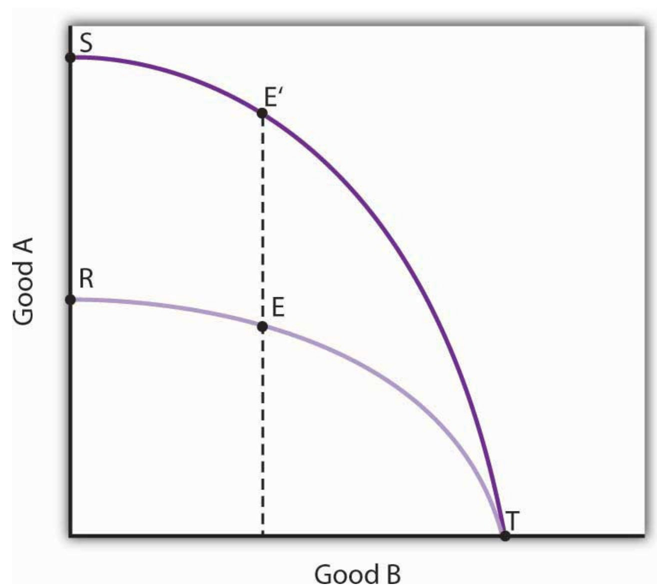
Increasing the quantity or quality of factors of production and/or improving technology will shift the production possibilities curve outward. This process is called economic growth. In the last 50 years, economic growth in the United States has resulted chiefly from increases in human capital and from technological advance.

Choices concerning the use of scarce resources take place within the context of a set of institutional arrangements that define an economic system. The principal distinctions between systems lie in the degree to which ownership of capital and natural resources and decision making authority over scarce resources are held by government or by private individuals. Economic systems include market capitalist, mixed, and command socialist economies. An increasing body of evidence suggests that market capitalist economies tend to be most productive; many command socialist and mixed economies are moving in the direction of market capitalist systems.

The presumption in favor of market-based systems does not preclude a role for government. Government is necessary to provide the system of laws on which market systems are founded. It may also be used to provide certain goods and services, to help individuals in need, and to regulate the actions of individuals and firms.

CONCEPT PROBLEMS

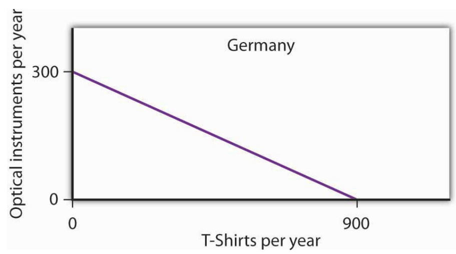
1. How does a college education increase one's human capital?
2. Why does the downward-sloping production possibilities curve imply that factors of production are scarce?
3. In what ways are the bowed-out shape of the production possibilities curve and the law of increasing opportunity cost related?
4. What is the relationship between the concept of comparative advantage and the law of increasing opportunity cost?
5. Suppose an economy can produce two goods, A and B. It is now operating at point E on production possibilities curve RT. An improvement in the technology available to produce good A shifts the curve to ST, and the economy selects point E′. How does this change affect the opportunity cost of producing an additional unit of good B?

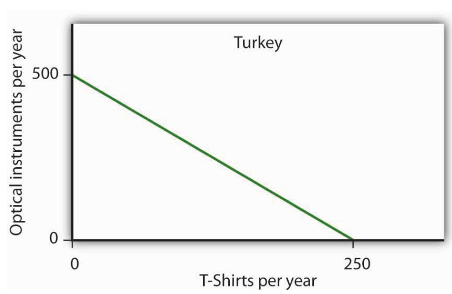


6. Could a nation's production possibilities curve ever shift inward? Explain what such a shift would mean, and discuss events that might cause such a shift to occur.
7. Suppose blue-eyed people were banned from working. How would this affect a nation's production possibilities curve?
8. Evaluate this statement: "The U.S. economy could achieve greater growth by devoting fewer resources to consumption and more to investment; it follows that such a shift would be desirable."
9. Two countries, Sportsland and Foodland, have similar total quantities of labor, capital, and natural resources. Both can produce two goods, figs and footballs. Sportsland's resources are particularly well suited to the production of footballs but are not very productive in producing figs. Foodland's resources are very productive when used for figs but are not capable of producing many footballs. In which country is the cost of additional footballs generally greater? Explain.
10. Suppose a country is committed to using its resources based on the reverse of comparative advantage doctrine: it first transfers those resources for which the cost is greatest, not lowest. Describe this country's production possibilities curve.
11. The U.S. Constitution bans states from restricting imports of goods and services from other states. Suppose this restriction did not exist and that states were allowed to limit imports of goods and services produced in other states. How do you think this would affect U.S. output? Explain.
12. By 1993, nations in the European Union (EU) had eliminated all barriers to the flow of goods, services, labor, and capital across their borders. Even such things as consumer protection laws and the types of plugs required to plug in appliances have been standardized to ensure that there will be no barriers to trade. How do you think this elimination of trade barriers affected EU output?
13. How did the technological changes described in the Case in Point "Technology Cuts Costs, Boosts Productivity and Profits" affect the production possibilities curve for the United States?

NUMERICAL PROBLEMS

1. Nathan can mow four lawns in a day or plant 20 trees in a day.
    a. Draw Nathan's production possibilities curve for mowing lawns and planting trees. Assume the production possibilities curve is linear and put the quantity of lawns mowed per day on the horizontal axis and the quantity of trees planted per day on the vertical axis.
    b. What is Nathan's opportunity cost of planting trees?
    c. What is Nathan's opportunity cost of mowing lawns?
2. David can mow four lawns in a day or plant four trees in a day.
    a. Draw David's production possibilities curve for mowing lawns and planting trees. Again, assume a linear production possibilities curve and put the quantity of lawns mowed per day on the horizontal axis.
    b. What is David's opportunity cost of planting trees?
    c. What is David's opportunity cost of mowing lawns?
3. Given the production information in problems 1 and 2 above, who has the comparative advantage in planting trees? Mowing lawns?
4. The exhibits below describe the production possibilities for Germany and Turkey.





    a. What is the slope of Germany's production possibilities curve?
    b. What is the slope of Turkey's production possibilities curve?
    c. What is the opportunity cost of producing T-shirts in Germany?
    d. What is the opportunity cost of producing T-shirts in Turkey?
    e. What is the opportunity cost of producing optical instruments in Germany?
    f. What is the opportunity cost of producing optical instruments in Turkey?
    g. In which good does Germany have a comparative advantage?
    h. In which good does Turkey have a comparative advantage?
5. The nation of Leisureland can produce two goods, bicycles and bowling balls. The western region of Leisureland can, if it devotes all its resources to bicycle production, produce 100 bicycles per month. Alternatively, it could devote all its resources to bowling balls and produce 400 per month—or it could produce any combination of bicycles and bowling balls lying on a straight line between these two extremes.
    a. Draw a production possibilities curve for western Leisureland (with bicycles on the vertical axis).
    b. What it is the opportunity cost of producing an additional bowling ball measured in terms of forgone bicycles in western Leisureland?
    c. Suppose that eastern Leisureland can, if it devotes all its resources to the production of bicycles, produce 400. If it devotes all its resources to bowling ball production, though, it can produce only 100. Draw the production possibilities curve for eastern Leisureland (again, assume it is linear and put bicycles on the vertical axis).

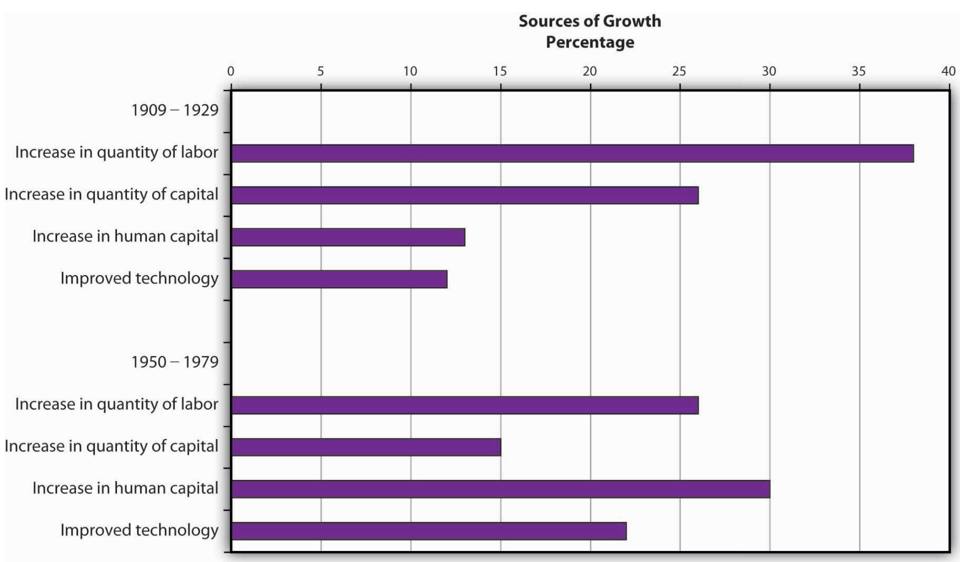
d. What is the opportunity cost of producing an additional bowling ball measured in terms of forgone bicycles in eastern Leisureland?
e. Explain the difference in opportunity cost between western and eastern Leisureland. Which region has a comparative advantage in producing bowling balls? Bicycles?
f. Draw the production possibilities curve for Leisureland, one that combines the curves for western and eastern Leisureland.
g. Suppose it is determined that 400 bicycles must be produced. How many bowling balls can be produced?
h. Where will these goods be produced?

6. The table below shows the production possibilities schedule for an economy.

Production Alternatives	Capital goods per period	Consumer goods per period
A	0	40
B	1	36
C	2	28
D	3	16
E	4	0

a. Putting capital goods per period on the horizontal axis and consumer goods per period on the vertical axis, graph the production possibilities curve for the economy.
b. If the economy is producing at alternative B, what is the opportunity cost to it of producing at alternative C instead?
c. If the economy is producing at alternative C, what is the opportunity cost to it of producing at alternative D instead?
d. Is it possible for this economy to produce 30 units of consumer goods per period while producing 1 unit of capital goods? Would this combination of goods represent efficient or inefficient production? Explain.
e. Which point, B or C, would lead to higher economic growth? Explain your answer.

7. The exhibit below shows the sources of growth in the United States between 1909 and 1929 and between 1950 and 1979, according to a study by Edward Denison.[1] (Note: The sources of economic growth are cumulative and, taken collectively, explain 100% of total growth over the period.)



a. Approximately what percentage of U.S. growth between 1909 and 1929 was due to increases in quantities of factors of production?
b. Approximately what percentage of U.S. growth between 1909 and 1929 was due to increases in quality of factors of production and technological improvement?

c. Approximately what percentage of U.S. growth between 1950 and 1979 was due to increases in quantities of factors of production?

d. Approximately what percentage of U.S. growth between 1950 and 1979 was due to increases in quality of factors of production and technological improvement?

ENDNOTES

1. Edward Denison, *The Sources of Economic Growth in the United States* (New York: Committee for Economic Development, 1962) and Edward Denison, *Trends in American Growth 1929–1982* (Washington, D.C.: Brookings Institutions, 1985).

CHAPTER 3

Demand and Supply

START UP: CRAZY FOR COFFEE

Starbucks Coffee Company revolutionized the coffee-drinking habits of millions of Americans. Starbucks, whose bright green-and-white logo is almost as familiar as the golden arches of McDonald's, began in Seattle in 1971. Fifteen years later it had grown into a chain of four stores in the Seattle area. Then in 1987 Howard Schultz, a former Starbucks employee, who had become enamored with the culture of Italian coffee bars during a trip to Italy, bought the company from its founders for $3.8 million. In 2008, Americans were willingly paying $3 or more for a cappuccino or a latté, and Starbuck's had grown to become an international chain, with over 16,000 stores around the world.

The change in American consumers' taste for coffee and the profits raked in by Starbucks lured other companies to get into the game. Retailers such as Seattle's Best Coffee and Gloria Jean's Coffees entered the market, and today there are thousands of coffee bars, carts, drive-throughs, and kiosks in downtowns, malls, and airports all around the country. Even McDonald's began selling specialty coffees.

But over the last decade the price of coffee beans has been quite volatile. Just as consumers were growing accustomed to their cappuccinos and lattés, in 1997, the price of coffee beans shot up. Excessive rain and labor strikes in coffee-growing areas of South America had reduced the supply of coffee, leading to a rise in its price. In the early 2000s, Vietnam flooded the market with coffee, and the price of coffee beans plummeted. More recently, weather conditions in various coffee-growing countries reduced supply, and the price of coffee beans went back up.

Markets, the institutions that bring together buyers and sellers, are always responding to events, such as bad harvests and changing consumer tastes that affect the prices and quantities of particular goods. The demand for some goods increases, while the demand for others decreases. The supply of some goods rises, while the supply of others falls. As such events unfold, prices adjust to keep markets in balance. This chapter explains how the market forces of demand and supply interact to determine equilibrium prices and equilibrium quantities of goods and services. We will see how prices and quantities adjust to changes in demand and supply and how changes in prices serve as signals to buyers and sellers.

markets

The institutions that bring together buyers and sellers.

The model of demand and supply that we shall develop in this chapter is one of the most powerful tools in all of economic analysis. You will be using it throughout your study of economics. We will first look at the variables that influence demand. Then we will turn to supply, and finally we will put demand and supply together to explore how the model of demand and supply operates. As we examine the model, bear in mind that demand is a representation of the behavior of buyers and that supply is a representation of the behavior of sellers. Buyers may be consumers purchasing groceries or producers purchasing iron ore to make steel. Sellers may be firms selling cars or households selling their labor services. We shall see that the ideas of demand and supply apply, whatever the identity of the buyers or sellers and whatever the good or service being exchanged in the market. In this chapter, we shall focus on buyers and sellers of goods and services.

1. DEMAND

LEARNING OBJECTIVES

1. **Define the quantity demanded of a good or service and illustrate it using a demand schedule and a demand curve.**
2. **Distinguish between the following pairs of concepts: demand and quantity demanded, demand schedule and demand curve, movement along and shift in a demand curve.**
3. **Identify demand shifters and determine whether a change in a demand shifter causes the demand curve to shift to the right or to the left.**

How many pizzas will people eat this year? How many doctor visits will people make? How many houses will people buy?

Each good or service has its own special characteristics that determine the quantity people are willing and able to consume. One is the price of the good or service itself. Other independent variables that are important determinants of demand include consumer preferences, prices of related goods and services, income, demographic characteristics such as population size, and buyer expectations. The number of pizzas people will purchase, for example, depends very much on whether they like pizza. It also depends on the prices for alternatives such as hamburgers or spaghetti. The number of doctor visits is likely to vary with income—people with higher incomes are likely to see a doctor more often than people with lower incomes. The demands for pizza, for doctor visits, and for housing are certainly affected by the age distribution of the population and its size.

While different variables play different roles in influencing the demands for different goods and services, economists pay special attention to one: the price of the good or service. Given the values of all the other variables that affect demand, a higher price tends to reduce the quantity people demand, and a lower price tends to increase it. A medium pizza typically sells for $5 to $10. Suppose the price were $30. Chances are, you would buy fewer pizzas at that price than you do now. Suppose pizzas typically sold for $2 each. At that price, people would be likely to buy more pizzas than they do now.

We will discuss first how price affects the quantity demanded of a good or service and then how other variables affect demand.

1.1 Price and the Demand Curve

Because people will purchase different quantities of a good or service at different prices, economists must be careful when speaking of the "demand" for something. They have therefore developed some specific terms for expressing the general concept of demand.

quantity demanded

The quantity buyers are willing and able to buy of a good or service at a particular price during a particular period, all other things unchanged.

The **quantity demanded** of a good or service is the quantity buyers are willing and able to buy at a particular price during a particular period, all other things unchanged. (As we learned, we can substitute the Latin phrase "ceteris paribus" for "all other things unchanged.") Suppose, for example, that 100,000 movie tickets are sold each month in a particular town at a price of $8 per ticket. That quantity—100,000—is the quantity of movie admissions demanded per month at a price of $8. If the price were $12, we would expect the quantity demanded to be less. If it were $4, we would expect the quantity demanded to be greater. The quantity demanded at each price would be different if other things that might affect it, such as the population of the town, were to change. That is why we add the qualifier that other things have not changed to the definition of quantity demanded.

demand schedule

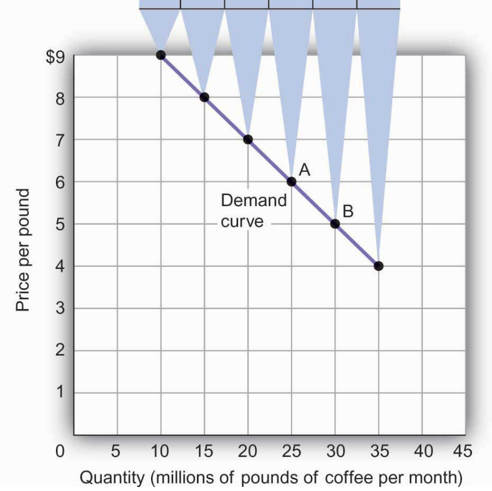
A table that shows the quantities of a good or service demanded at different prices during a particular period, all other things unchanged.

A **demand schedule** is a table that shows the quantities of a good or service demanded at different prices during a particular period, all other things unchanged. To introduce the concept of a demand schedule, let us consider the demand for coffee in the United States. We will ignore differences among types of coffee beans and roasts, and speak simply of coffee. The table in Figure 3.1 shows quantities of coffee that will be demanded each month at prices ranging from $9 to $4 per pound; the table is a demand schedule. We see that the higher the price, the lower the quantity demanded.

FIGURE 3.1 A Demand Schedule and a Demand Curve

The table is a demand schedule; it shows quantities of coffee demanded per month in the United States at particular prices, all other things unchanged. These data are then plotted on the demand curve. At point A on the curve, 25 million pounds of coffee per month are demanded at a price of $6 per pound. At point B, 30 million pounds of coffee per month are demanded at a price of $5 per pound.

Price per pound ($)	9	8	7	6	5	4
Quantity demanded per month (millions of pounds)	10	15	20	25	30	35



The information given in a demand schedule can be presented with a **demand curve**, which is a graphical representation of a demand schedule. A demand curve thus shows the relationship between the price and quantity demanded of a good or service during a particular period, all other things unchanged. The demand curve in Figure 3.1 shows the prices and quantities of coffee demanded that are given in the demand schedule. At point A, for example, we see that 25 million pounds of coffee per month are demanded at a price of $6 per pound. By convention, economists graph price on the vertical axis and quantity on the horizontal axis.

demand curve

A graphical representation of a demand schedule.

Price alone does not determine the quantity of coffee or any other good that people buy. To isolate the effect of changes in price on the quantity of a good or service demanded, however, we show the quantity demanded at each price, assuming that those other variables remain unchanged. We do the same thing in drawing a graph of the relationship between any two variables; we assume that the values of other variables that may affect the variables shown in the graph (such as income or population) remain unchanged for the period under consideration.

A change in price, with no change in any of the other variables that affect demand, results in a movement *along* the demand curve. For example, if the price of coffee falls from $6 to $5 per pound, consumption rises from 25 million pounds to 30 million pounds per month. That is a movement from point A to point B along the demand curve in Figure 3.1. A movement along a demand curve that results from a change in price is called a **change in quantity demanded**. Note that a change in quantity demanded is not a change or shift in the demand curve; it is a movement *along* the demand curve.

change in quantity demanded

A movement along a demand curve that results from a change in price.

The negative slope of the demand curve in Figure 3.1 suggests a key behavioral relationship of economics. All other things unchanged, the **law of demand** holds that, for virtually all goods and services, a higher price leads to a reduction in quantity demanded and a lower price leads to an increase in quantity demanded.

law of demand

For virtually all goods and services, a higher price leads to a reduction in quantity demanded and a lower price leads to an increase in quantity demanded.

The law of demand is called a law because the results of countless studies are consistent with it. Undoubtedly, you have observed one manifestation of the law. When a store finds itself with an overstock of some item, such as running shoes or tomatoes, and needs to sell these items quickly, what does it do? It typically has a sale, expecting that a lower price will increase the quantity demanded. In general, we expect the law of demand to hold. Given the values of other variables that influence demand, a higher price reduces the quantity demanded. A lower price increases the quantity demanded. Demand curves, in short, slope downward.

1.2 Changes in Demand

change in demand

A shift in a demand curve.

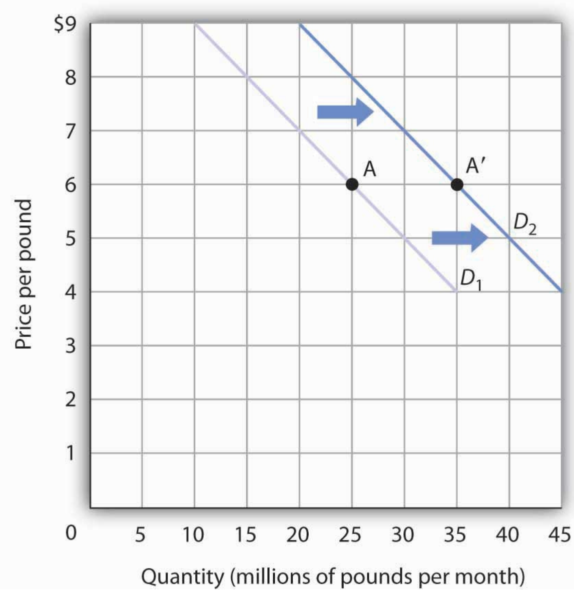
Of course, price alone does not determine the quantity of a good or service that people consume. Coffee consumption, for example, will be affected by such variables as income and population. Preferences also play a role. The story at the beginning of the chapter illustrates as much. Starbucks "turned people on" to coffee. We also expect other prices to affect coffee consumption. People often eat doughnuts or bagels with their coffee, so a reduction in the price of doughnuts or bagels might induce people to drink more coffee. An alternative to coffee is tea, so a reduction in the price of tea might result in the consumption of more tea and less coffee. Thus, a change in any one of the variables held constant in constructing a demand schedule will change the quantities demanded at each price. The result will be a *shift* in the entire demand curve rather than a movement along the demand curve. A *shift* in a demand curve is called a **change in demand**.

Suppose, for example, that something happens to increase the quantity of coffee demanded at each price. Several events could produce such a change: an increase in incomes, an increase in population, or an increase in the price of tea would each be likely to increase the quantity of coffee demanded at each price. Any such change produces a new demand schedule. Figure 3.2 shows such a change in the demand schedule for coffee. We see that the quantity of coffee demanded per month is greater at each price than before. We show that graphically as a shift in the demand curve. The original curve, labeled D_1, shifts to the right to D_2. At a price of $6 per pound, for example, the quantity demanded rises from 25 million pounds per month (point A) to 35 million pounds per month (point A′).

FIGURE 3.2 An Increase in Demand

An increase in the quantity of a good or service demanded at each price is shown as an increase in demand. Here, the original demand curve D_1 shifts to D_2. Point A on D_1 corresponds to a price of $6 per pound and a quantity demanded of 25 million pounds of coffee per month. On the new demand curve D_2, the quantity demanded at this price rises to 35 million pounds of coffee per month (point A′).

Price	Old quantity demanded	New quantity demanded
$9	10	20
8	15	25
7	20	30
6	25	35
5	30	40
4	35	45



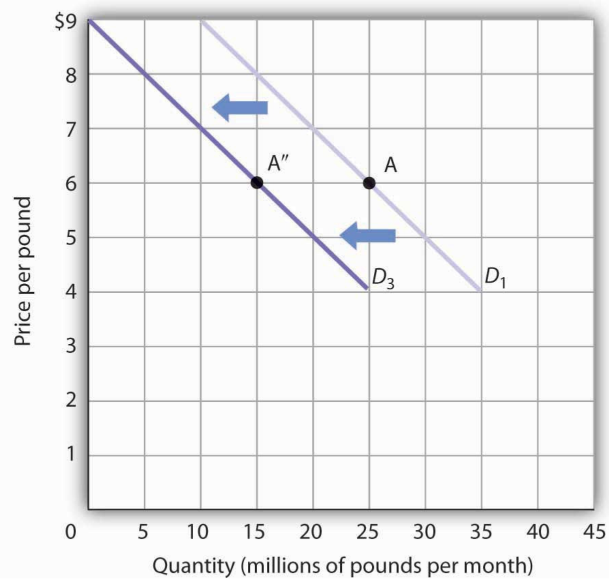
Just as demand can increase, it can decrease. In the case of coffee, demand might fall as a result of events such as a reduction in population, a reduction in the price of tea, or a change in preferences. For example, a definitive finding that the caffeine in coffee contributes to heart disease, which is currently being debated in the scientific community, could change preferences and reduce the demand for coffee.

A reduction in the demand for coffee is illustrated in Figure 3.3. The demand schedule shows that less coffee is demanded at each price than in Figure 3.1. The result is a shift in demand from the original curve D_1 to D_3. The quantity of coffee demanded at a price of $6 per pound falls from 25 million pounds per month (point A) to 15 million pounds per month (point A″). Note, again, that a change in quantity demanded, ceteris paribus, refers to a movement *along* the demand curve, while a change in demand refers to a *shift* in the demand curve.

FIGURE 3.3 A Reduction in Demand

A reduction in demand occurs when the quantities of a good or service demanded fall at each price. Here, the demand schedule shows a lower quantity of coffee demanded at each price than we had in Figure 3.1. The reduction shifts the demand curve for coffee to D_3 from D_1. The quantity demanded at a price of \$6 per pound, for example, falls from 25 million pounds per month (point A) to 15 million pounds of coffee per month (point A″).

Price	Old quantity demanded	New quantity demanded
\$9	10	0
8	15	5
7	20	10
6	25	15
5	30	20
4	35	25



A variable that can change the quantity of a good or service demanded at each price is called a **demand shifter**. When these other variables change, the all-other-things-unchanged conditions behind the original demand curve no longer hold. Although different goods and services will have different demand shifters, the demand shifters are likely to include (1) consumer preferences, (2) the prices of related goods and services, (3) income, (4) demographic characteristics, and (5) buyer expectations. Next we look at each of these.

demand shifter

A variable that can change the quantity of a good or service demanded at each price.

Preferences

Changes in preferences of buyers can have important consequences for demand. We have already seen how Starbucks supposedly increased the demand for coffee. Another example is reduced demand for cigarettes caused by concern about the effect of smoking on health. A change in preferences that makes one good or service more popular will shift the demand curve to the right. A change that makes it less popular will shift the demand curve to the left.

Prices of Related Goods and Services

Suppose the price of doughnuts were to fall. Many people who drink coffee enjoy dunking doughnuts in their coffee; the lower price of doughnuts might therefore increase the demand for coffee, shifting the demand curve for coffee to the right. A lower price for tea, however, would be likely to reduce coffee demand, shifting the demand curve for coffee to the left.

In general, if a reduction in the price of one good increases the demand for another, the two goods are called **complements**. If a reduction in the price of one good reduces the demand for another, the two goods are called **substitutes**. These definitions hold in reverse as well: two goods are complements if an increase in the price of one reduces the demand for the other, and they are substitutes if an increase in the price of one increases the demand for the other. Doughnuts and coffee are complements; tea and coffee are substitutes.

Complementary goods are goods used in conjunction with one another. Tennis rackets and tennis balls, eggs and bacon, and stationery and postage stamps are complementary goods. Substitute goods are goods used instead of one another. iPODs, for example, are likely to be substitutes for CD players. Breakfast cereal is a substitute for eggs. A file attachment to an e-mail is a substitute for both a fax machine and postage stamps.

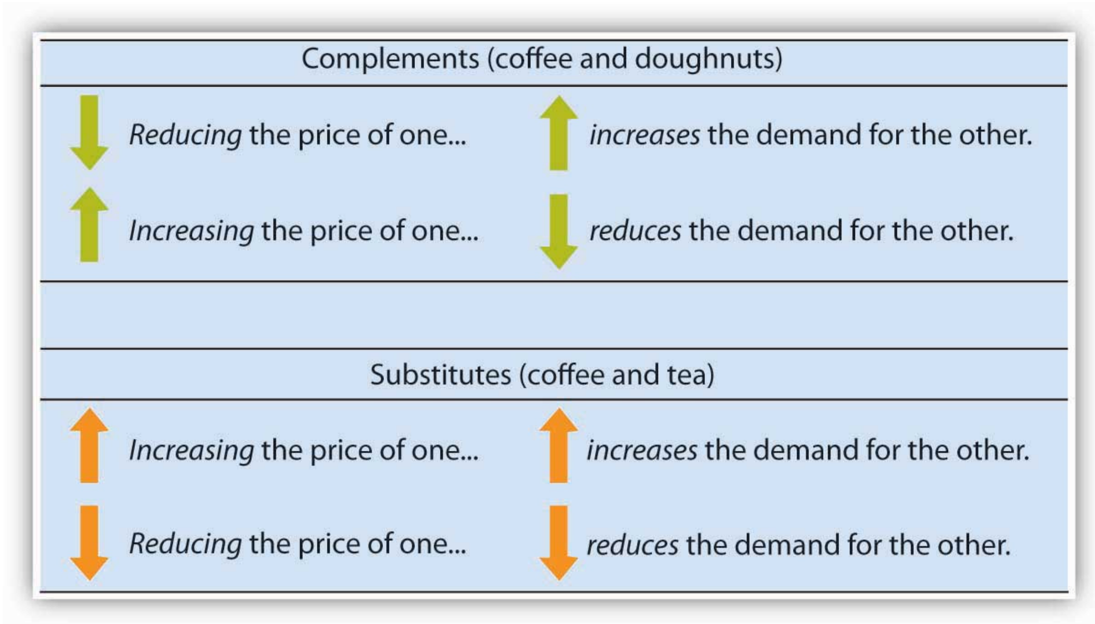
complements

Two goods for which an increase in price of one reduces the demand for the other.

substitutes

Two goods for which an increase in price of one increases the demand for the other.

FIGURE 3.4



Income

As incomes rise, people increase their consumption of many goods and services, and as incomes fall, their consumption of these goods and services falls. For example, an increase in income is likely to raise the demand for gasoline, ski trips, new cars, and jewelry. There are, however, goods and services for which consumption falls as income rises—and rises as income falls. As incomes rise, for example, people tend to consume more fresh fruit but less canned fruit.

A good for which demand increases when income increases is called a **normal good**. A good for which demand decreases when income increases is called an **inferior good**. An increase in income shifts the demand curve for fresh fruit (a normal good) to the right; it shifts the demand curve for canned fruit (an inferior good) to the left.

normal good

A good for which demand increases when income increases.

inferior good

A good for which demand decreases when income increases.

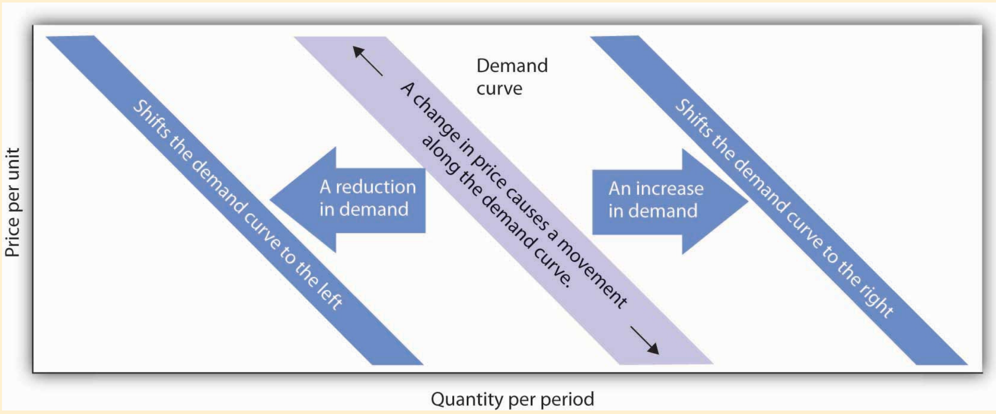
Demographic Characteristics

The number of buyers affects the total quantity of a good or service that will be bought; in general, the greater the population, the greater the demand. Other demographic characteristics can affect demand as well. As the share of the population over age 65 increases, the demand for medical services, ocean cruises, and motor homes increases. The birth rate in the United States fell sharply between 1955 and 1975 but has gradually increased since then. That increase has raised the demand for such things as infant supplies, elementary school teachers, soccer coaches, in-line skates, and college education. Demand can thus shift as a result of changes in both the number and characteristics of buyers.

Buyer Expectations

The consumption of goods that can be easily stored, or whose consumption can be postponed, is strongly affected by buyer expectations. The expectation of newer TV technologies, such as high-definition TV, could slow down sales of regular TVs. If people expect gasoline prices to rise tomorrow, they will fill up their tanks today to try to beat the price increase. The same will be true for goods such as automobiles and washing machines: an expectation of higher prices in the future will lead to more purchases today. If the price of a good is expected to fall, however, people are likely to reduce their purchases today and await tomorrow's lower prices. The expectation that computer prices will fall, for example, can reduce current demand.

Heads Up!



It is crucial to distinguish between a change in quantity demanded, which is a movement along the demand curve caused by a change in price, and a change in demand, which implies a shift of the demand curve itself. A change in demand is caused by a change in a demand shifter. An increase in demand is a shift of the demand curve to the right. A decrease in demand is a shift in the demand curve to the left. This drawing of a demand curve highlights the difference.

KEY TAKEAWAYS

- The quantity demanded of a good or service is the quantity buyers are willing and able to buy at a particular price during a particular period, all other things unchanged.
- A demand schedule is a table that shows the quantities of a good or service demanded at different prices during a particular period, all other things unchanged.
- A demand curve shows graphically the quantities of a good or service demanded at different prices during a particular period, all other things unchanged.
- All other things unchanged, the law of demand holds that, for virtually all goods and services, a higher price induces a reduction in quantity demanded and a lower price induces an increase in quantity demanded.
- A change in the price of a good or service causes a change in the quantity demanded—a movement *along* the demand curve.
- A change in a demand shifter causes a change in demand, which is shown as a *shift* of the demand curve. Demand shifters include preferences, the prices of related goods and services, income, demographic characteristics, and buyer expectations.
- Two goods are substitutes if an increase in the price of one causes an increase in the demand for the other. Two goods are complements if an increase in the price of one causes a decrease in the demand for the other.
- A good is a normal good if an increase in income causes an increase in demand. A good is an inferior good if an increase in income causes a decrease in demand.

TRY IT!

All other things unchanged, what happens to the demand curve for DVD rentals if there is (a) an increase in the price of movie theater tickets, (b) a decrease in family income, or (c) an increase in the price of DVD rentals? In answering this and other "Try It!" problems in this chapter, draw and carefully label a set of axes. On the horizontal axis of your graph, show the quantity of DVD rentals. It is necessary to specify the time period to which your quantity pertains (e.g., "per period," "per week," or "per year"). On the vertical axis show the price per DVD rental. Since you do not have specific data on prices and quantities demanded, make a "free-hand" drawing of the curve or curves you are asked to examine. Focus on the general shape and position of the curve(s) before and after events occur. Draw new curve(s) to show what happens in each of the circumstances given. The curves could shift to the left or to the right, or stay where they are.

Case in Point: Solving Campus Parking Problems Without Adding More Parking Spaces

© 2010 Jupiterimages Corporation

Unless you attend a "virtual" campus, chances are you have engaged in more than one conversation about how hard it is to find a place to park on campus. Indeed, according to Clark Kerr, a former president of the University of California system, a university is best understood as a group of people "held together by a common grievance over parking."

Clearly, the demand for campus parking spaces has grown substantially over the past few decades. In surveys conducted by Daniel Kenney, Ricardo Dumont, and Ginger Kenney, who work for the campus design company Sasaki and Associates, it was found that 7 out of 10 students own their own cars. They have interviewed "many students who confessed to driving from their dormitories to classes that were a five-minute walk away," and they argue that the deterioration of college environments is largely attributable to the increased use of cars on campus and that colleges could better service their missions by not adding more parking spaces.

Since few universities charge enough for parking to even cover the cost of building and maintaining parking lots, the rest is paid for by all students as part of tuition. Their research shows that "for every 1,000 parking spaces, the median institution loses almost $400,000 a year for surface parking, and more than $1,200,000 for structural parking." Fear of a backlash from students and their parents, as well as from faculty and staff, seems to explain why campus administrators do not simply raise the price of parking on campus.

While Kenney and his colleagues do advocate raising parking fees, if not all at once then over time, they also suggest some subtler, and perhaps politically more palatable, measures—in particular, shifting the demand for parking spaces to the left by lowering the prices of substitutes.

Two examples they noted were at the University of Washington and the University of Colorado at Boulder. At the University of Washington, car poolers may park for free. This innovation has reduced purchases of single-occupancy parking permits by 32% over a decade. According to University of Washington assistant director of transportation services Peter Dewey, "Without vigorously managing our parking and providing commuter alternatives, the university would have been faced with adding approximately 3,600 parking spaces, at a cost of over $100 million…The university has created opportunities to make capital investments in buildings supporting education instead of structures for cars." At the University of Colorado, free public transit has increased use of buses and light rail from 300,000 to 2 million trips per year over the last decade. The increased use of mass transit has allowed the university to avoid constructing nearly 2,000 parking spaces, which has saved about $3.6 million annually.

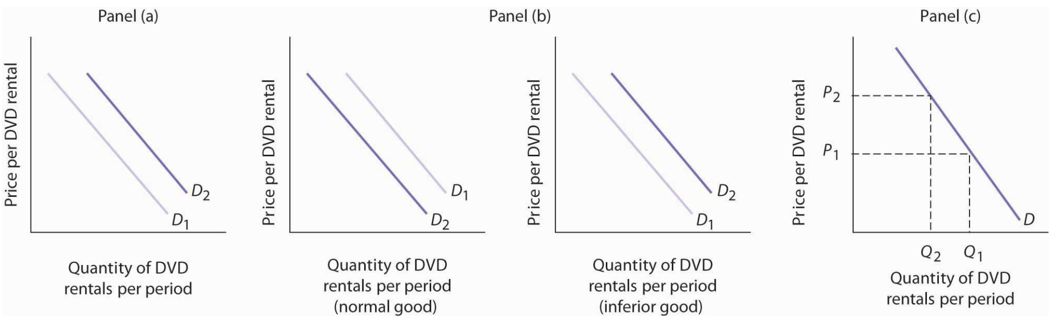
Sources: Daniel R. Kenney, "How to Solve Campus Parking Problems Without Adding More Parking," The Chronicle of Higher Education, March 26, 2004, Section B, pp. B22-B23.

ANSWER TO TRY IT! PROBLEM

Since going to the movies is a substitute for watching a DVD at home, an increase in the price of going to the movies should cause more people to switch from going to the movies to staying at home and renting DVDs. Thus, the demand curve for DVD rentals will shift to the right when the price of movie theater tickets increases [Panel (a)].

A decrease in family income will cause the demand curve to shift to the left if DVD rentals are a normal good but to the right if DVD rentals are an inferior good. The latter may be the case for some families, since staying at home and watching DVDs is a cheaper form of entertainment than taking the family to the movies. For most others, however, DVD rentals are probably a normal good [Panel (b)].

An increase in the price of DVD rentals does not shift the demand curve for DVD rentals at all; rather, an increase in price, say from P_1 to P_2, is a movement upward to the left along the demand curve. At a higher price, people will rent fewer DVDs, say Q_2 instead of Q_1, ceteris paribus [Panel (c)].



2. SUPPLY

LEARNING OBJECTIVES

1. **Define the quantity supplied of a good or service and illustrate it using a supply schedule and a supply curve.**
2. **Distinguish between the following pairs of concepts: supply and quantity supplied, supply schedule and supply curve, movement along and shift in a supply curve.**
3. **Identify supply shifters and determine whether a change in a supply shifter causes the supply curve to shift to the right or to the left.**

What determines the quantity of a good or service sellers are willing to offer for sale? Price is one factor; ceteris paribus, a higher price is likely to induce sellers to offer a greater quantity of a good or service. Production cost is another determinant of supply. Variables that affect production cost include the prices of factors used to produce the good or service, returns from alternative activities, technology, the expectations of sellers, and natural events such as weather changes. Still another factor affecting the quantity of a good that will be offered for sale is the number of sellers—the greater the number of sellers of a particular good or service, the greater will be the quantity offered at any price per time period.

2.1 Price and the Supply Curve

The **quantity supplied** of a good or service is the quantity sellers are willing to sell at a particular price during a particular period, all other things unchanged. Ceteris paribus, the receipt of a higher price increases profits and induces sellers to increase the quantity they supply.

quantity supplied

The quantity sellers are willing to sell of a good or service at a particular price during a particular period, all other things unchanged.

In general, when there are many sellers of a good, an increase in price results in an increase in quantity supplied, and this relationship is often referred to as the law of supply. We will see, though, through our exploration of microeconomics, that there are a number of exceptions to this relationship. There are cases in which a higher price will not induce an increase in quantity supplied. Goods that cannot be produced, such as additional land on the corner of Park Avenue and 56th Street in Manhattan, are fixed in supply—a higher price cannot induce an increase in the quantity supplied. There are even cases, which we investigate in microeconomic analysis, in which a higher price induces a reduction in the quantity supplied.

supply schedule

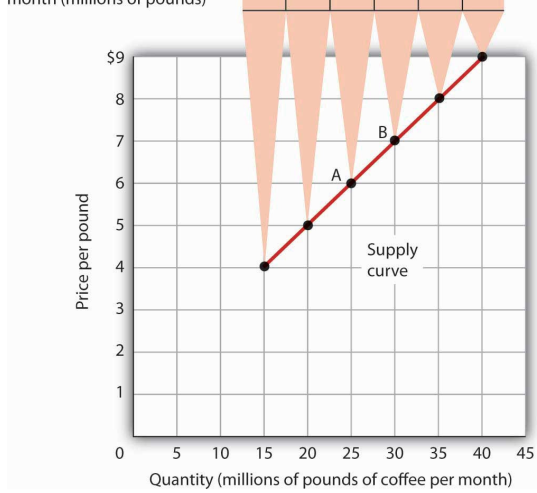
A table that shows quantities supplied at different prices during a particular period, all other things unchanged.

Generally speaking, however, when there are many sellers of a good, an increase in price results in a greater quantity supplied. The relationship between price and quantity supplied is suggested in a **supply schedule**, a table that shows quantities supplied at different prices during a particular period, all other things unchanged. Figure 3.8 gives a supply schedule for the quantities of coffee that will be supplied per month at various prices, ceteris paribus. At a price of $4 per pound, for example, producers are willing to supply 15 million pounds of coffee per month. A higher price, say $6 per pound, induces sellers to supply a greater quantity—25 million pounds of coffee per month.

FIGURE 3.8 A Supply Schedule and a Supply Curve

The supply schedule shows the quantity of coffee that will be supplied in the United States each month at particular prices, all other things unchanged. The same information is given graphically in the supply curve. The values given here suggest a positive relationship between price and quantity supplied.

Price per pound ($)	4	5	6	7	8	9
Quantity supplied per month (millions of pounds)	15	20	25	30	35	40



supply curve

A graphical representation of a supply schedule.

A **supply curve** is a graphical representation of a supply schedule. It shows the relationship between price and quantity supplied during a particular period, all other things unchanged. Because the relationship between price and quantity supplied is generally positive, supply curves are generally upward sloping. The supply curve for coffee in Figure 3.8 shows graphically the values given in the supply schedule.

change in quantity supplied

Movement along the supply curve caused by a change in price.

A change in price causes a movement *along* the supply curve; such a movement is called a **change in quantity supplied**. As is the case with a change in quantity demanded, a change in quantity supplied does not shift the supply curve. By definition, it is a movement along the supply curve. For example, if the price rises from $6 per pound to $7 per pound, the quantity supplied rises from 25 million pounds per month to 30 million pounds per month. That's a movement from point A to point B along the supply curve in Figure 3.8.

2.2 Changes in Supply

change in supply

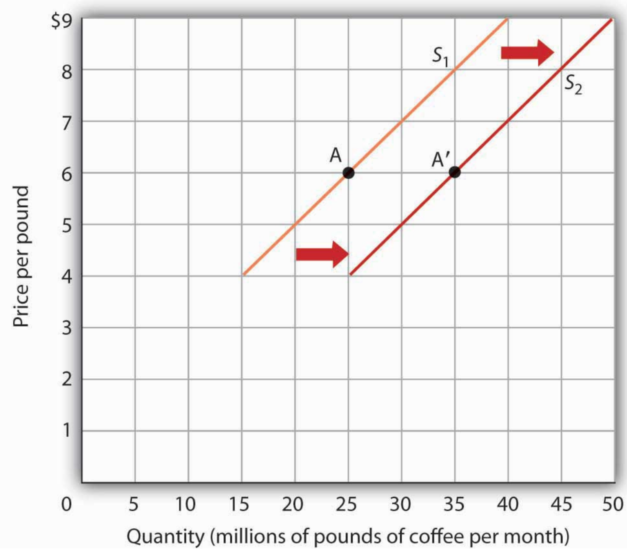
A shift in the supply curve.

When we draw a supply curve, we assume that other variables that affect the willingness of sellers to supply a good or service are unchanged. It follows that a change in any of those variables will cause a **change in supply**, which is a shift in the supply curve. A change that increases the quantity of a good or service supplied at each price shifts the supply curve to the right. Suppose, for example, that the price of fertilizer falls. That will reduce the cost of producing coffee and thus increase the quantity of coffee producers will offer for sale at each price. The supply schedule in Figure 3.9 shows an increase in the quantity of coffee supplied at each price. We show that increase graphically as a shift in the supply curve from S_1 to S_2. We see that the quantity supplied at each price increases by 10 million pounds of coffee per month. At point A on the original supply curve S_1, for example, 25 million pounds of coffee per month are supplied at a price of $6 per pound. After the increase in supply, 35 million pounds per month are supplied at the same price (point A′ on curve S_2).

FIGURE 3.9 An Increase in Supply

If there is a change in supply that increases the quantity supplied at each price, as is the case in the supply schedule here, the supply curve shifts to the right. At a price of $6 per pound, for example, the quantity supplied rises from the previous level of 25 million pounds per month on supply curve S_1 (point A) to 35 million pounds per month on supply curve S_2 (point A′).

Price	Old quantity supplied	New quantity supplied
$4	15	25
5	20	30
6	25	35
7	30	40
8	35	45
9	40	50

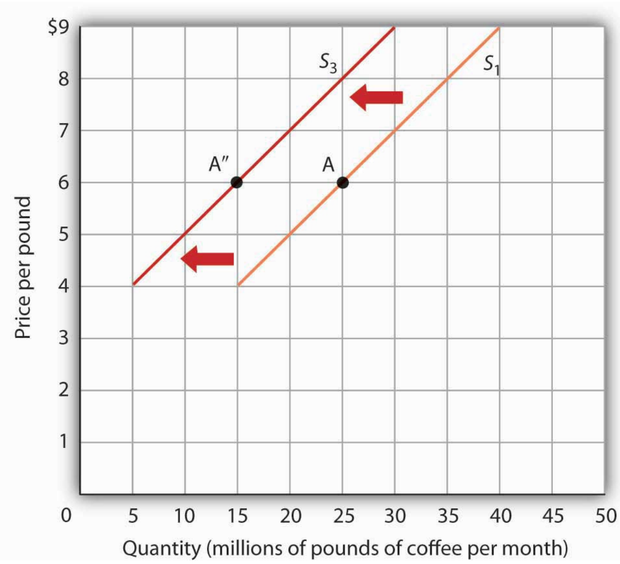


An event that reduces the quantity supplied at each price shifts the supply curve to the left. An increase in production costs and excessive rain that reduces the yields from coffee plants are examples of events that might reduce supply. Figure 3.10 shows a reduction in the supply of coffee. We see in the supply schedule that the quantity of coffee supplied falls by 10 million pounds of coffee per month at each price. The supply curve thus shifts from S_1 to S_3.

FIGURE 3.10 A Reduction in Supply

A change in supply that reduces the quantity supplied at each price shifts the supply curve to the left. At a price of $6 per pound, for example, the original quantity supplied was 25 million pounds of coffee per month (point A). With a new supply curve S_3, the quantity supplied at that price falls to 15 million pounds of coffee per month (point A″).

Price	Old quantity supplied	New quantity supplied
$4	15	5
5	20	10
6	25	15
7	30	20
8	35	25
9	40	30



supply shifter

A variable that can change the quantity of a good or service supplied at each price.

A variable that can change the quantity of a good or service supplied at each price is called a **supply shifter**. Supply shifters include (1) prices of factors of production, (2) returns from alternative activities, (3) technology, (4) seller expectations, (5) natural events, and (6) the number of sellers. When these other variables change, the all-other-things-unchanged conditions behind the original supply curve no longer hold. Let us look at each of the supply shifters.

Prices of Factors of Production

A change in the price of labor or some other factor of production will change the cost of producing any given quantity of the good or service. This change in the cost of production will change the quantity that suppliers are willing to offer at any price. An increase in factor prices should decrease the quantity suppliers will offer at any price, shifting the supply curve to the left. A reduction in factor prices increases the quantity suppliers will offer at any price, shifting the supply curve to the right.

Suppose coffee growers must pay a higher wage to the workers they hire to harvest coffee or must pay more for fertilizer. Such increases in production cost will cause them to produce a smaller quantity at each price, shifting the supply curve for coffee to the left. A reduction in any of these costs increases supply, shifting the supply curve to the right.

Returns from Alternative Activities

To produce one good or service means forgoing the production of another. The concept of opportunity cost in economics suggests that the value of the activity forgone is the opportunity cost of the activity chosen; this cost should affect supply. For example, one opportunity cost of producing eggs is not selling chickens. An increase in the price people are willing to pay for fresh chicken would make it more profitable to sell chickens and would thus increase the opportunity cost of producing eggs. It would shift the supply curve for eggs to the left, reflecting a decrease in supply.

Technology

A change in technology alters the combinations of inputs or the types of inputs required in the production process. An improvement in technology usually means that fewer and/or less costly inputs are needed. If the cost of production is lower, the profits available at a given price will increase, and producers will produce more. With more produced at every price, the supply curve will shift to the right, meaning an increase in supply.

Impressive technological changes have occurred in the computer industry in recent years. Computers are much smaller and are far more powerful than they were only a few years ago—and they are much cheaper to produce. The result has been a huge increase in the supply of computers, shifting the supply curve to the right.

While we usually think of technology as enhancing production, declines in production due to problems in technology are also possible. Outlawing the use of certain equipment without pollution-control devices has increased the cost of production for many goods and services, thereby reducing profits available at any price and shifting these supply curves to the left.

Seller Expectations

All supply curves are based in part on seller expectations about future market conditions. Many decisions about production and selling are typically made long before a product is ready for sale. Those decisions necessarily depend on expectations. Changes in seller expectations can have important effects on price and quantity.

Consider, for example, the owners of oil deposits. Oil pumped out of the ground and used today will be unavailable in the future. If a change in the international political climate leads many owners to expect that oil prices will rise in the future, they may decide to leave their oil in the ground, planning to sell it later when the price is higher. Thus, there will be a decrease in supply; the supply curve for oil will shift to the left.

Natural Events

Storms, insect infestations, and drought affect agricultural production and thus the supply of agricultural goods. If something destroys a substantial part of an agricultural crop, the supply curve will shift to the left. The terrible cyclone that killed more than 50,000 people in Myanmar in 2008 also destroyed some of the country's prime rice growing land. That shifted the supply curve for rice to the left. If there is an unusually good harvest, the supply curve will shift to the right.

The Number of Sellers

The supply curve for an industry, such as coffee, includes all the sellers in the industry. A change in the number of sellers in an industry changes the quantity available at each price and thus changes supply. An increase in the number of sellers supplying a good or service shifts the supply curve to the right; a reduction in the number of sellers shifts the supply curve to the left.

The market for cellular phone service has been affected by an increase in the number of firms offering the service. Over the past decade, new cellular phone companies emerged, shifting the supply curve for cellular phone service to the right.

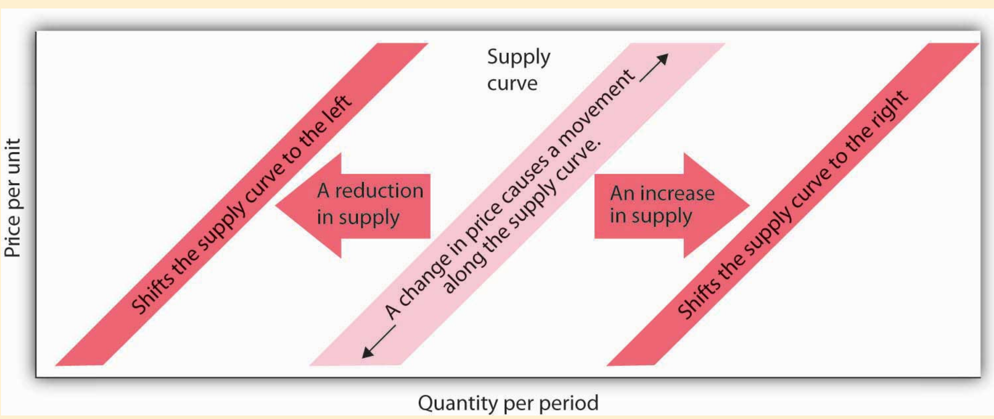
Heads Up!

There are two special things to note about supply curves. The first is similar to the Heads Up! on demand curves: it is important to distinguish carefully between changes in supply and changes in quantity supplied. A change in supply results from a change in a supply shifter and implies a shift of the supply curve to the right or left. A change in price produces a change in quantity supplied and induces a movement along the supply curve. A change in price does not shift the supply curve.

The second caution relates to the interpretation of increases and decreases in supply. Notice that in Figure 3.9 an increase in supply is shown as a shift of the supply curve to the right; the curve shifts in the direction of increasing quantity with respect to the horizontal axis. In Figure 3.10 a reduction in supply is shown as a shift of the supply curve to the left; the curve shifts in the direction of decreasing quantity with respect to the horizontal axis.

Because the supply curve is upward sloping, a shift to the right produces a new curve that in a sense lies "below" the original curve. Students sometimes make the mistake of thinking of such a shift as a shift "down" and therefore as a reduction in supply. Similarly, it is easy to make the mistake of showing an increase in supply with a new curve that lies "above" the original curve. But that is a reduction in supply!

To avoid such errors, focus on the fact that an increase in supply is an increase in the quantity supplied at each price and shifts the supply curve in the direction of increased quantity on the horizontal axis. Similarly, a reduction in supply is a reduction in the quantity supplied at each price and shifts the supply curve in the direction of a lower quantity on the horizontal axis.



KEY TAKEAWAYS

- The quantity supplied of a good or service is the quantity sellers are willing to sell at a particular price during a particular period, all other things unchanged.
- A supply schedule shows the quantities supplied at different prices during a particular period, all other things unchanged. A supply curve shows this same information graphically.
- A change in the price of a good or service causes a change in the quantity supplied—a movement *along* the supply curve.
- A change in a supply shifter causes a change in supply, which is shown as a *shift* of the supply curve. Supply shifters include prices of factors of production, returns from alternative activities, technology, seller expectations, natural events, and the number of sellers.
- An increase in supply is shown as a shift to the right of a supply curve; a decrease in supply is shown as a shift to the left.

TRY IT!

If all other things are unchanged, what happens to the supply curve for DVD rentals if there is (a) an increase in wages paid to DVD rental store clerks, (b) an increase in the price of DVD rentals, or (c) an increase in the number of DVD rental stores? Draw a graph that shows what happens to the supply curve in each circumstance. The supply curve can shift to the left or to the right, or stay where it is. Remember to label the axes and curves, and remember to specify the time period (e.g., "DVDs rented per week").

Case in Point: The Monks of St. Benedict's Get Out of the Egg Business

It was cookies that lured the monks of St. Benedict's out of the egg business, and now private retreat sponsorship is luring them away from cookies.

St. Benedict's is a Benedictine monastery, nestled on a ranch high in the Colorado Rockies, about 20 miles down the road from Aspen. The monastery's 15 monks operate the ranch to support themselves and to provide help for poor people in the area. They lease out about 3,500 acres of their land to cattle and sheep grazers, produce cookies, and sponsor private retreats. They used to produce eggs.

Attracted by potential profits and the peaceful nature of the work, the monks went into the egg business in 1967. They had 10,000 chickens producing their Monastery Eggs brand. For a while, business was good. Very good. Then, in the late 1970s, the price of chicken feed started to rise rapidly.

"When we started in the business, we were paying \$60 to \$80 a ton for feed—delivered," recalls the monastery's abbot, Father Joseph Boyle. "By the late 1970s, our cost had more than doubled. We were paying \$160 to \$200 a ton. That really hurt, because feed represents a large part of the cost of producing eggs."

The monks adjusted to the blow. "When grain prices were lower, we'd pull a hen off for a few weeks to molt, then return her to laying. After grain prices went up, it was 12 months of laying and into the soup pot," Father Joseph says.

Grain prices continued to rise in the 1980s and increased the costs of production for all egg producers. It caused the supply of eggs to fall. Demand fell at the same time, as Americans worried about the cholesterol in eggs. Times got tougher in the egg business.

"We were still making money in the financial sense," Father Joseph says. "But we tried an experiment in 1985 producing cookies, and it was a success. We finally decided that devoting our time and energy to the cookies would pay off better than the egg business, so we quit the egg business in 1986."

The mail-order cookie business was good to the monks. They sold 200,000 ounces of Monastery Cookies in 1987.

By 1998, however, they had limited their production of cookies, selling only locally and to gift shops. Since 2000, they have switched to "providing private retreats for individuals and groups—about 40 people per month," according to Brother Charles.

The monks' calculation of their opportunity costs revealed that they would earn a higher return through sponsorship of private retreats than in either cookies or eggs. This projection has proved correct.

And there is another advantage as well.

"The chickens didn't stop laying eggs on Sunday," Father Joseph chuckles. "When we shifted to cookies we could take Sundays off. We weren't hemmed in the way we were with the chickens." The move to providing retreats is even better in this regard. Since guests provide their own meals, most of the monastery's effort goes into planning and scheduling, which frees up even more of their time for other worldly as well as spiritual pursuits.

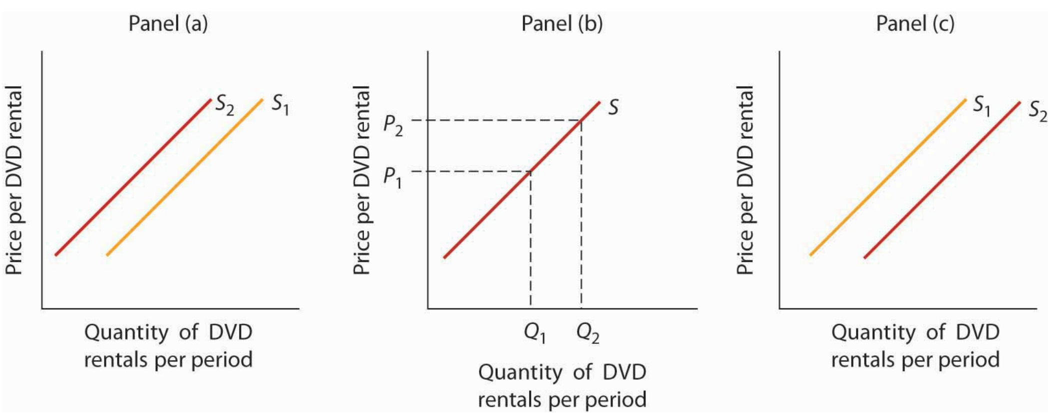
Source: Personal interviews.

ANSWER TO TRY IT! PROBLEM

DVD rental store clerks are a factor of production in the DVD rental market. An increase in their wages raises the cost of production, thereby causing the supply curve of DVD rentals to shift to the left [Panel (a)]. (*Caution*: It is possible that you thought of the wage increase as an increase in income, a demand shifter, that would lead to an increase in demand, but this would be incorrect. The question refers only to wages of DVD rental store clerks. They may rent some DVD, but their impact on total demand would be negligible. Besides, we have no information on what has happened overall to incomes of people who rent DVDs. We do know, however, that the cost of a factor of production, which is a supply shifter, increased.)

An increase in the price of DVD rentals does not shift the supply curve at all; rather, it corresponds to a movement upward to the right along the supply curve. At a higher price of P_2 instead of P_1, a greater quantity of DVD rentals, say Q_2 instead of Q_1, will be supplied [Panel (b)].

An increase in the number of stores renting DVDs will cause the supply curve to shift to the right [Panel (c)].



3. DEMAND, SUPPLY, AND EQUILIBRIUM

LEARNING OBJECTIVES

1. Use demand and supply to explain how equilibrium price and quantity are determined in a market.
2. Understand the concepts of surpluses and shortages and the pressures on price they generate.
3. Explain the impact of a change in demand or supply on equilibrium price and quantity.
4. Explain how the circular flow model provides an overview of demand and supply in product and factor markets and how the model suggests ways in which these markets are linked.

model of demand and supply

Model that uses demand and supply curves to explain the determination of price and quantity in a market.

In this section we combine the demand and supply curves we have just studied into a new model. The **model of demand and supply** uses demand and supply curves to explain the determination of price and quantity in a market.

3.1 The Determination of Price and Quantity

The logic of the model of demand and supply is simple. The demand curve shows the quantities of a particular good or service that buyers will be willing and able to purchase at each price during a specified period. The supply curve shows the quantities that sellers will offer for sale at each price during that same period. By putting the two curves together, we should be able to find a price at which the quantity buyers are willing and able to purchase equals the quantity sellers will offer for sale.

equilibrium price

The price at which quantity demanded equals quantity supplied.

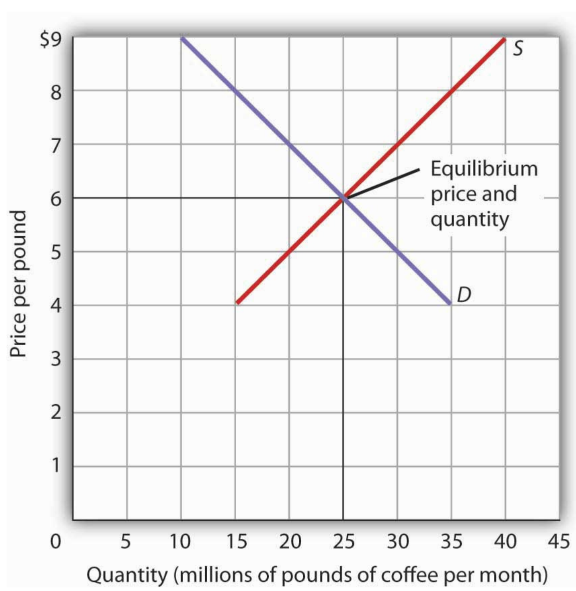
equilibrium quantity

The quantity demanded and supplied at the equilibrium price.

Figure 3.14 combines the demand and supply data introduced in Figure 3.1 and Figure 3.8 Notice that the two curves intersect at a price of $6 per pound—at this price the quantities demanded and supplied are equal. Buyers want to purchase, and sellers are willing to offer for sale, 25 million pounds of coffee per month. The market for coffee is in equilibrium. Unless the demand or supply curve shifts, there will be no tendency for price to change. The **equilibrium price** in any market is the price at which quantity demanded equals quantity supplied. The equilibrium price in the market for coffee is thus $6 per pound. The **equilibrium quantity** is the quantity demanded and supplied at the equilibrium price.

FIGURE 3.14 The Determination of Equilibrium Price and Quantity

When we combine the demand and supply curves for a good in a single graph, the point at which they intersect identifies the equilibrium price and equilibrium quantity. Here, the equilibrium price is $6 per pound. Consumers demand, and suppliers supply, 25 million pounds of coffee per month at this price.



With an upward-sloping supply curve and a downward-sloping demand curve, there is only a single price at which the two curves intersect. This means there is only one price at which equilibrium is achieved. It follows that at any price other than the equilibrium price, the market will not be in equilibrium. We next examine what happens at prices other than the equilibrium price.

Surpluses

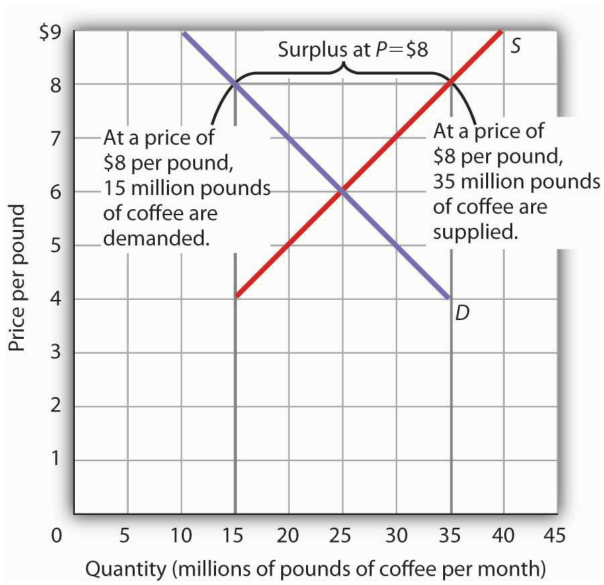
Figure 3.15 shows the same demand and supply curves we have just examined, but this time the initial price is $8 per pound of coffee. Because we no longer have a balance between quantity demanded and quantity supplied, this price is not the equilibrium price. At a price of $8, we read over to the demand curve to determine the quantity of coffee consumers will be willing to buy—15 million pounds per month. The supply curve tells us what sellers will offer for sale—35 million pounds per month. The difference, 20 million pounds of coffee per month, is called a surplus. More generally, a **surplus** is the amount by which the quantity supplied exceeds the quantity demanded at the current price. There is, of course, no surplus at the equilibrium price; a surplus occurs only if the current price exceeds the equilibrium price.

surplus

The amount by which the quantity supplied exceeds the quantity demanded at the current price.

FIGURE 3.15 A Surplus in the Market for Coffee

At a price of $8, the quantity supplied is 35 million pounds of coffee per month and the quantity demanded is 15 million pounds per month; there is a surplus of 20 million pounds of coffee per month. Given a surplus, the price will fall quickly toward the equilibrium level of $6.



A surplus in the market for coffee will not last long. With unsold coffee on the market, sellers will begin to reduce their prices to clear out unsold coffee. As the price of coffee begins to fall, the quantity of coffee supplied begins to decline. At the same time, the quantity of coffee demanded begins to rise. Remember that the reduction in quantity supplied is a movement *along* the supply curve—the curve itself does not shift in response to a reduction in price. Similarly, the increase in quantity demanded is a movement *along* the demand curve—the demand curve does not shift in response to a reduction in price. Price will continue to fall until it reaches its equilibrium level, at which the demand and supply curves intersect. At that point, there will be no tendency for price to fall further. In general, surpluses in the marketplace are short-lived. The prices of most goods and services adjust quickly, eliminating the surplus. Later on, we will discuss some markets in which adjustment of price to equilibrium may occur only very slowly or not at all.

Shortages

shortage

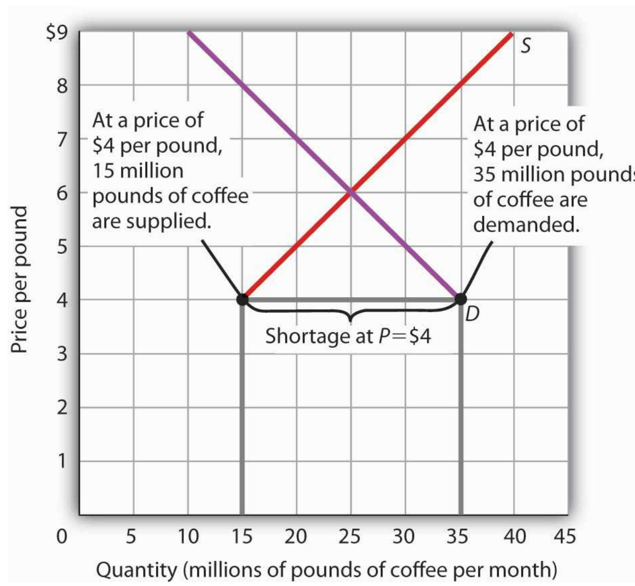
The amount by which the quantity demanded exceeds the quantity supplied at the current price.

Just as a price above the equilibrium price will cause a surplus, a price below equilibrium will cause a shortage. A **shortage** is the amount by which the quantity demanded exceeds the quantity supplied at the current price.

Figure 3.16 shows a shortage in the market for coffee. Suppose the price is $4 per pound. At that price, 15 million pounds of coffee would be supplied per month, and 35 million pounds would be demanded per month. When more coffee is demanded than supplied, there is a shortage.

FIGURE 3.16 A Shortage in the Market for Coffee

At a price of $4 per pound, the quantity of coffee demanded is 35 million pounds per month and the quantity supplied is 15 million pounds per month. The result is a shortage of 20 million pounds of coffee per month.

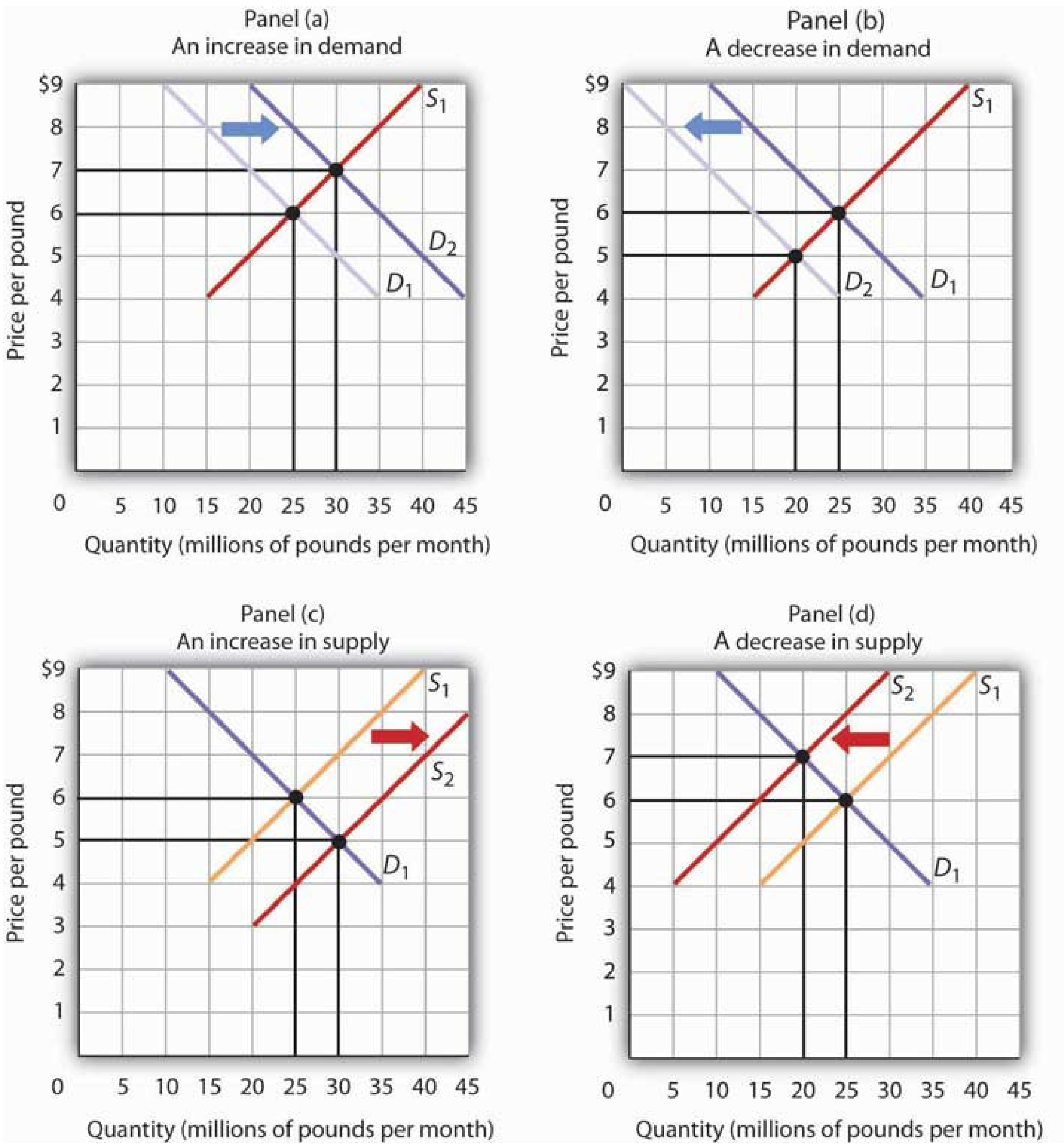


In the face of a shortage, sellers are likely to begin to raise their prices. As the price rises, there will be an increase in the quantity supplied (but not a change in supply) and a reduction in the quantity demanded (but not a change in demand) until the equilibrium price is achieved.

3.2 Shifts in Demand and Supply

FIGURE 3.17 Changes in Demand and Supply

A change in demand or in supply changes the equilibrium solution in the model. Panels (a) and (b) show an increase and a decrease in demand, respectively; Panels (c) and (d) show an increase and a decrease in supply, respectively.



A change in one of the variables (shifters) held constant in any model of demand and supply will create a change in demand or supply. A shift in a demand or supply curve changes the equilibrium price and equilibrium quantity for a good or service. Figure 3.17 combines the information about changes in the demand and supply of coffee presented in Figure 3.2 Figure 3.3 Figure 3.9 and Figure 3.10 In each case, the original equilibrium price is $6 per pound, and the corresponding equilibrium quantity is 25 million pounds of coffee per month. Figure 3.17 shows what happens with an increase in demand, a reduction in demand, an increase in supply, and a reduction in supply. We then look at what happens if both curves shift simultaneously. Each of these possibilities is discussed in turn below.

An Increase in Demand

An increase in demand for coffee shifts the demand curve to the right, as shown in Panel (a) of Figure 3.17. The equilibrium price rises to $7 per pound. As the price rises to the new equilibrium level, the quantity supplied increases to 30 million pounds of coffee per month. Notice that the supply curve does not shift; rather, there is a movement along the supply curve.

Demand shifters that could cause an increase in demand include a shift in preferences that leads to greater coffee consumption; a lower price for a complement to coffee, such as doughnuts; a higher price for a substitute for coffee, such as tea; an increase in income; and an increase in population. A change in buyer expectations, perhaps due to predictions of bad weather lowering expected yields on coffee plants and increasing future coffee prices, could also increase current demand.

A Decrease in Demand

Panel (b) of Figure 3.17 shows that a decrease in demand shifts the demand curve to the left. The equilibrium price falls to $5 per pound. As the price falls to the new equilibrium level, the quantity supplied decreases to 20 million pounds of coffee per month.

Demand shifters that could reduce the demand for coffee include a shift in preferences that makes people want to consume less coffee; an increase in the price of a complement, such as doughnuts; a reduction in the price of a substitute, such as tea; a reduction in income; a reduction in population; and a change in buyer expectations that leads people to expect lower prices for coffee in the future.

An Increase in Supply

An increase in the supply of coffee shifts the supply curve to the right, as shown in Panel (c) of Figure 3.17. The equilibrium price falls to $5 per pound. As the price falls to the new equilibrium level, the quantity of coffee demanded increases to 30 million pounds of coffee per month. Notice that the demand curve does not shift; rather, there is movement along the demand curve.

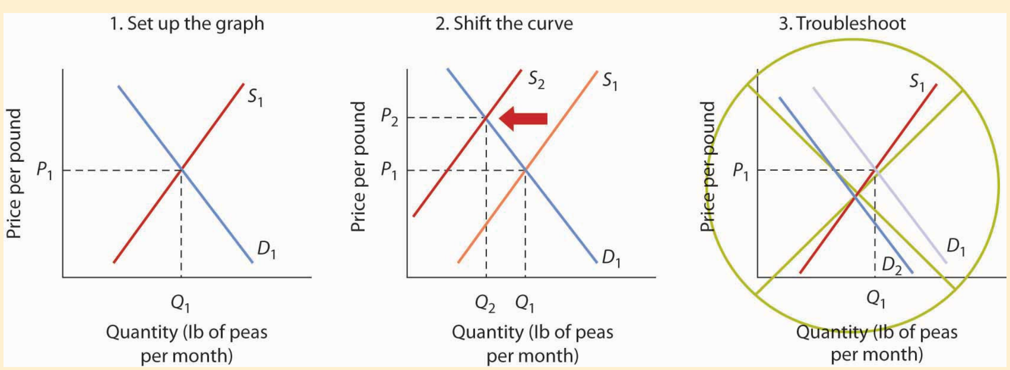
Possible supply shifters that could increase supply include a reduction in the price of an input such as labor, a decline in the returns available from alternative uses of the inputs that produce coffee, an improvement in the technology of coffee production, good weather, and an increase in the number of coffee-producing firms.

A Decrease in Supply

Panel (d) of Figure 3.17 shows that a decrease in supply shifts the supply curve to the left. The equilibrium price rises to $7 per pound. As the price rises to the new equilibrium level, the quantity demanded decreases to 20 million pounds of coffee per month.

Possible supply shifters that could reduce supply include an increase in the prices of inputs used in the production of coffee, an increase in the returns available from alternative uses of these inputs, a decline in production because of problems in technology (perhaps caused by a restriction on pesticides used to protect coffee beans), a reduction in the number of coffee-producing firms, or a natural event, such as excessive rain.

Heads Up!



You are likely to be given problems in which you will have to shift a demand or supply curve.

Suppose you are told that an invasion of pod-crunching insects has gobbled up half the crop of fresh peas, and you are asked to use demand and supply analysis to predict what will happen to the price and quantity of peas demanded and supplied. Here are some suggestions.

Put the quantity of the good you are asked to analyze on the horizontal axis and its price on the vertical axis. Draw a downward-sloping line for demand and an upward-sloping line for supply. The initial equilibrium price is determined by the intersection of the two curves. Label the equilibrium solution. You may find it helpful to use a number for the equilibrium price instead of the letter "P." Pick a price that seems plausible, say, 79¢ per pound. Do not worry about the precise positions of the demand and supply curves; you cannot be expected to know what they are.

Step 2 can be the most difficult step; the problem is to decide which curve to shift. The key is to remember the difference between a change in demand or supply and a change in quantity demanded or supplied. At each price, ask yourself whether the given event would change the quantity demanded. Would the fact that a bug has attacked the pea crop change the quantity demanded at a price of, say, 79¢ per pound? Clearly not; none of the demand shifters have changed. The event would, however, reduce the quantity supplied at this price, and the supply curve would shift to the left. There is a change in supply and a reduction in the quantity demanded. There is no change in demand.

Next check to see whether the result you have obtained makes sense. The graph in Step 2 makes sense; it shows price rising and quantity demanded falling.

It is easy to make a mistake such as the one shown in the third figure of this Heads Up! One might, for example, reason that when fewer peas are available, fewer will be demanded, and therefore the demand curve will shift to the left. This suggests the price of peas will fall—but that does not make sense. If only half as many fresh peas were available, their price would surely rise. The error here lies in confusing a change in quantity demanded with a change in demand. Yes, buyers will end up buying fewer peas. But no, they will not demand fewer peas at each price than before; the demand curve does not shift.

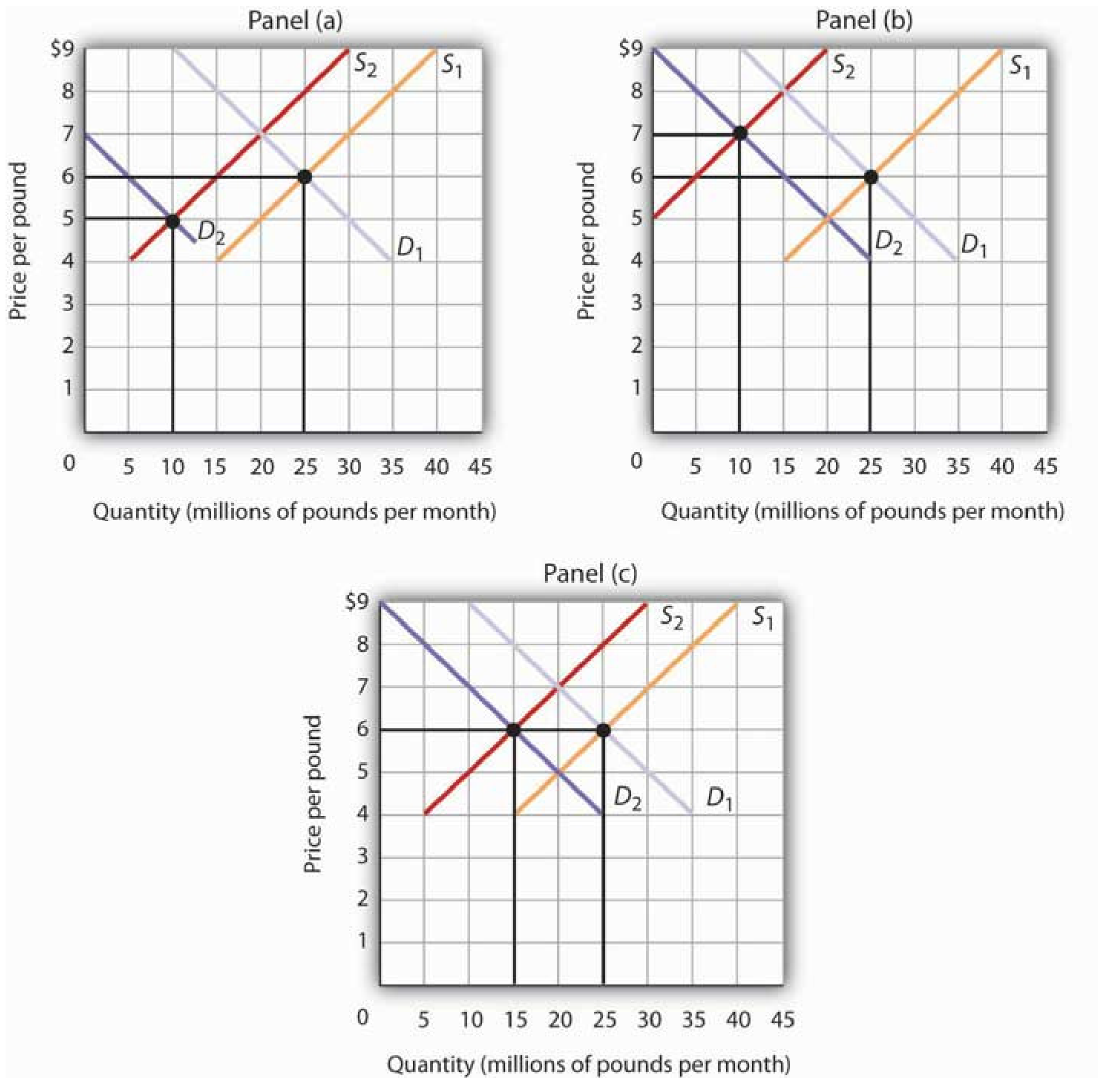
Simultaneous Shifts

As we have seen, when *either* the demand or the supply curve shifts, the results are unambiguous; that is, we know what will happen to both equilibrium price and equilibrium quantity, so long as we know whether demand or supply increased or decreased. However, in practice, several events may occur at around the same time that cause *both* the demand and supply curves to shift. To figure out what happens to equilibrium price and equilibrium quantity, we must know not only in which direction the demand and supply curves have shifted but also the relative amount by which each curve shifts. Of course, the demand and supply curves could shift in the same direction or in opposite directions, depending on the specific events causing them to shift.

For example, all three panels of Figure 3.19 show a decrease in demand for coffee (caused perhaps by a decrease in the price of a substitute good, such as tea) and a simultaneous decrease in the supply of coffee (caused perhaps by bad weather). Since reductions in demand and supply, considered separately, each cause the equilibrium quantity to fall, the impact of both curves shifting simultaneously to the left means that the new equilibrium quantity of coffee is less than the old equilibrium quantity. The effect on the equilibrium price, though, is ambiguous. Whether the equilibrium price is higher, lower, or unchanged depends on the extent to which each curve shifts.

FIGURE 3.19 Simultaneous Decreases in Demand and Supply

Both the demand and the supply of coffee decrease. Since decreases in demand and supply, considered separately, each cause equilibrium quantity to fall, the impact of both decreasing simultaneously means that a new equilibrium quantity of coffee must be less than the old equilibrium quantity. In Panel (a), the demand curve shifts farther to the left than does the supply curve, so equilibrium price falls. In Panel (b), the supply curve shifts farther to the left than does the demand curve, so the equilibrium price rises. In Panel (c), both curves shift to the left by the same amount, so equilibrium price stays the same.

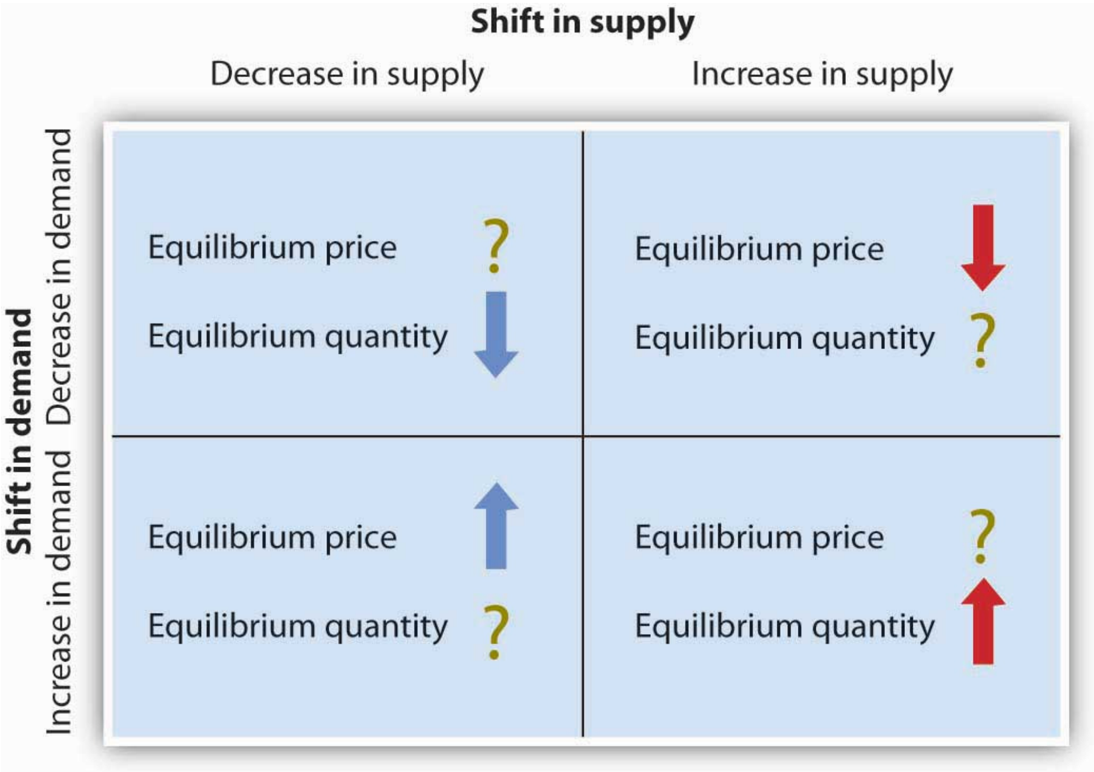


If the demand curve shifts farther to the left than does the supply curve, as shown in Panel (a) of Figure 3.19, then the equilibrium price will be lower than it was before the curves shifted. In this case the new equilibrium price falls from \$6 per pound to \$5 per pound. If the shift to the left of the supply curve is greater than that of the demand curve, the equilibrium price will be higher than it was before, as shown in Panel (b). In this case, the new equilibrium price rises to \$7 per pound. In Panel (c), since both curves shift to the left by the same amount, equilibrium price does not change; it remains \$6 per pound.

Regardless of the scenario, changes in equilibrium price and equilibrium quantity resulting from two different events need to be considered separately. If both events cause equilibrium price or quantity to move in the same direction, then clearly price or quantity can be expected to move in that direction. If one event causes price or quantity to rise while the other causes it to fall, the extent by which each curve shifts is critical to figuring out what happens. Figure 3.20 summarizes what may happen to equilibrium price and quantity when demand and supply both shift.

FIGURE 3.20 Simultaneous Shifts in Demand and Supply

If simultaneous shifts in demand and supply cause equilibrium price or quantity to move in the same direction, then equilibrium price or quantity clearly moves in that direction. If the shift in one of the curves causes equilibrium price or quantity to rise while the shift in the other curve causes equilibrium price or quantity to fall, then the relative amount by which each curve shifts is critical to figuring out what happens to that variable.



As demand and supply curves shift, prices adjust to maintain a balance between the quantity of a good demanded and the quantity supplied. If prices did not adjust, this balance could not be maintained.

Notice that the demand and supply curves that we have examined in this chapter have all been drawn as linear. This simplification of the real world makes the graphs a bit easier to read without sacrificing the essential point: whether the curves are linear or nonlinear, demand curves are downward sloping and supply curves are generally upward sloping. As circumstances that shift the demand curve or the supply curve change, we can analyze what will happen to price and what will happen to quantity.

3.3 An Overview of Demand and Supply: The Circular Flow Model

Implicit in the concepts of demand and supply is a constant interaction and adjustment that economists illustrate with the circular flow model. The **circular flow model** provides a look at how markets work and how they are related to each other. It shows flows of spending and income through the economy.

circular flow model

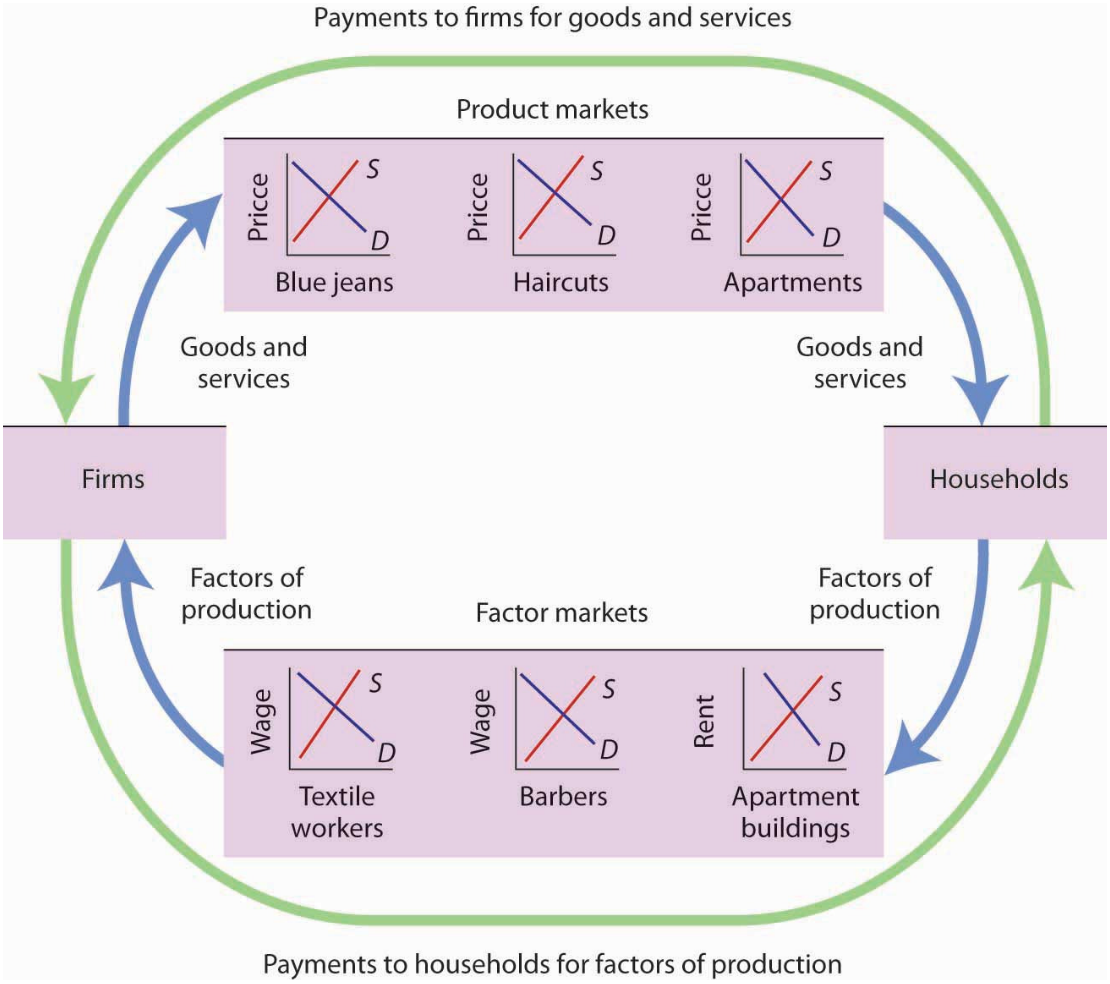
Model that provides a look at how markets work and how they are related to each other.

A great deal of economic activity can be thought of as a process of exchange between households and firms. Firms supply goods and services to households. Households buy these goods and services from firms. Households supply factors of production—labor, capital, and natural resources—that firms require. The payments firms make in exchange for these factors represent the incomes households earn.

The flow of goods and services, factors of production, and the payments they generate is illustrated in Figure 3.21. This circular flow model of the economy shows the interaction of households and firms as they exchange goods and services and factors of production. For simplicity, the model here shows only the private domestic economy; it omits the government and foreign sectors.

FIGURE 3.21 **The Circular Flow of Economic Activity**

This simplified circular flow model shows flows of spending between households and firms through product and factor markets. The inner arrows show goods and services flowing from firms to households and factors of production flowing from households to firms. The outer flows show the payments for goods, services, and factors of production. These flows, in turn, represent millions of individual markets for products and factors of production.



product markets

Markets in which firms supply goods and services demanded by households.

factor markets

Markets in which households supply factors of production—labor, capital, and natural resources—demanded by firms.

The circular flow model shows that goods and services that households demand are supplied by firms in **product markets**. The exchange for goods and services is shown in the top half of Figure 3.21. The bottom half of the exhibit illustrates the exchanges that take place in factor markets. **factor markets** are markets in which households supply factors of production—labor, capital, and natural resources—demanded by firms.

Our model is called a circular flow model because households use the income they receive from their supply of factors of production to buy goods and services from firms. Firms, in turn, use the payments they receive from households to pay for their factors of production.

The demand and supply model developed in this chapter gives us a basic tool for understanding what is happening in each of these product or factor markets and also allows us to see how these markets are interrelated. In Figure 3.21, markets for three goods and services that households want—blue jeans, haircuts, and apartments—create demands by firms for textile workers, barbers, and apartment buildings. The equilibrium of supply and demand in each market determines the price and quantity of that item. Moreover, a change in equilibrium in one market will affect equilibrium in related markets. For example, an increase in the demand for haircuts would lead to an increase in demand for barbers. Equilibrium price and quantity could rise in both markets. For some purposes, it will be adequate to simply look at a single market, whereas at other times we will want to look at what happens in related markets as well.

In either case, the model of demand and supply is one of the most widely used tools of economic analysis. That widespread use is no accident. The model yields results that are, in fact, broadly consistent with what we observe in the marketplace. Your mastery of this model will pay big dividends in your study of economics.

KEY TAKEAWAYS

- The equilibrium price is the price at which the quantity demanded equals the quantity supplied. It is determined by the intersection of the demand and supply curves.
- A surplus exists if the quantity of a good or service supplied exceeds the quantity demanded at the current price; it causes downward pressure on price. A shortage exists if the quantity of a good or service demanded exceeds the quantity supplied at the current price; it causes upward pressure on price.
- An increase in demand, all other things unchanged, will cause the equilibrium price to rise; quantity supplied will increase. A decrease in demand will cause the equilibrium price to fall; quantity supplied will decrease.
- An increase in supply, all other things unchanged, will cause the equilibrium price to fall; quantity demanded will increase. A decrease in supply will cause the equilibrium price to rise; quantity demanded will decrease.
- To determine what happens to equilibrium price and equilibrium quantity when both the supply and demand curves shift, you must know in which direction each of the curves shifts and the extent to which each curve shifts.
- The circular flow model provides an overview of demand and supply in product and factor markets and suggests how these markets are linked to one another.

TRY IT!

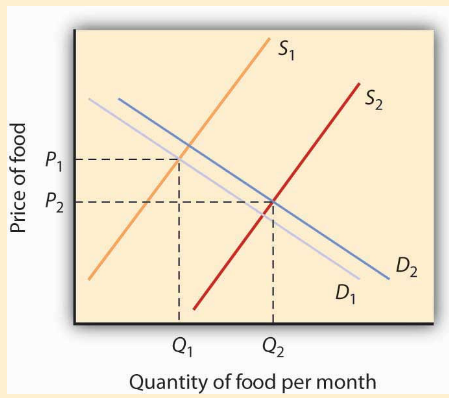
What happens to the equilibrium price and the equilibrium quantity of DVD rentals if the price of movie theater tickets increases and wages paid to DVD rental store clerks increase, all other things unchanged? Be sure to show all possible scenarios, as was done in Figure 3.19. Again, you do not need actual numbers to arrive at an answer. Just focus on the general position of the curve(s) before and after events occurred.

Case in Point: Demand, Supply, and Obesity

Why are so many Americans fat? Put so crudely, the question may seem rude, but, indeed, the number of obese Americans has increased by more than 50% over the last generation, and obesity may now be the nation's number one health problem. According to Sturm Roland in a recent RAND Corporation study, "Obesity appears to have a stronger association with the occurrence of chronic medical conditions, reduced physical health-related quality of life and increased health care and medication expenditures than smoking or problem drinking."

Many explanations of rising obesity suggest higher demand for food. What more apt picture of our sedentary life style is there than spending the afternoon watching a ballgame on TV, while eating chips and salsa, followed by a dinner of a lavishly topped, take-out pizza? Higher income has also undoubtedly contributed to a rightward shift in the demand curve for food. Plus, any additional food intake translates into more weight increase because we spend so few calories preparing it, either directly or in the process of earning the income to buy it. A study by economists Darius Lakdawalla and Tomas Philipson suggests that about 60% of the recent growth in weight may be explained in this way—that is, demand has shifted to the right, leading to an increase in the equilibrium quantity of food consumed and, given our less strenuous life styles, even more weight gain than can be explained simply by the increased amount we are eating.

What accounts for the remaining 40% of the weight gain? Lakdawalla and Philipson further reason that a rightward shift in demand would by itself lead to an increase in the quantity of food as well as an increase in the price of food. The problem they have with this explanation is that over the post-World War II period, the relative price of food has declined by an average of 0.2 percentage points per year. They explain the fall in the price of food by arguing that agricultural innovation has led to a substantial rightward shift in the supply curve of food. As shown, lower food prices and a higher equilibrium quantity of food have resulted from simultaneous rightward shifts in demand and supply and that the rightward shift in the supply of food from S_1 to S_2 has been substantially larger than the rightward shift in the demand curve from D_1 to D_2.



Sources: Roland, Sturm, "The Effects of Obesity, Smoking, and Problem Drinking on Chronic Medical Problems and Health Care Costs," Health Affairs, 2002; 21(2): 245–253. Lakdawalla, Darius and Tomas Philipson, "The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination," National Bureau of Economic Research Working Paper no. w8946, May 2002.

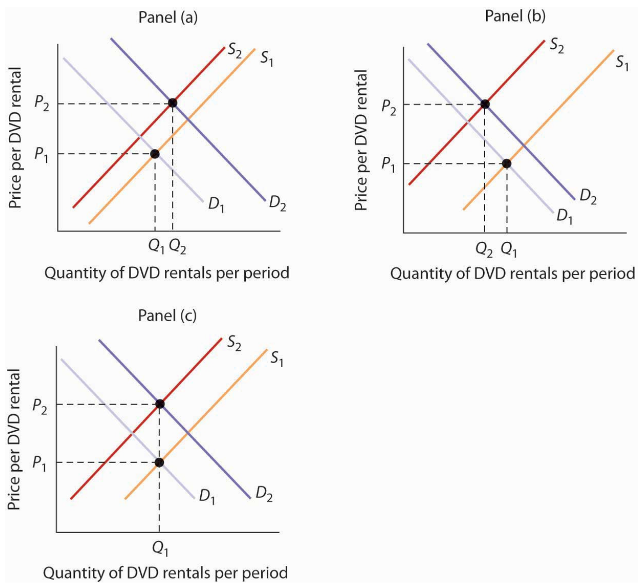
ANSWER TO TRY IT! PROBLEM

An increase in the price of movie theater tickets (a substitute for DVD rentals) will cause the demand curve for DVD rentals to shift to the right. An increase in the wages paid to DVD rental store clerks (an increase in the cost of a factor of production) shifts the supply curve to the left. Each event taken separately causes equilibrium price to rise. Whether equilibrium quantity will be higher or lower depends on which curve shifted more.

If the demand curve shifted more, then the equilibrium quantity of DVD rentals will rise [Panel (a)].

If the supply curve shifted more, then the equilibrium quantity of DVD rentals will fall [Panel (b)].

If the curves shifted by the same amount, then the equilibrium quantity of DVD rentals would not change [Panel (c)].



4. REVIEW AND PRACTICE

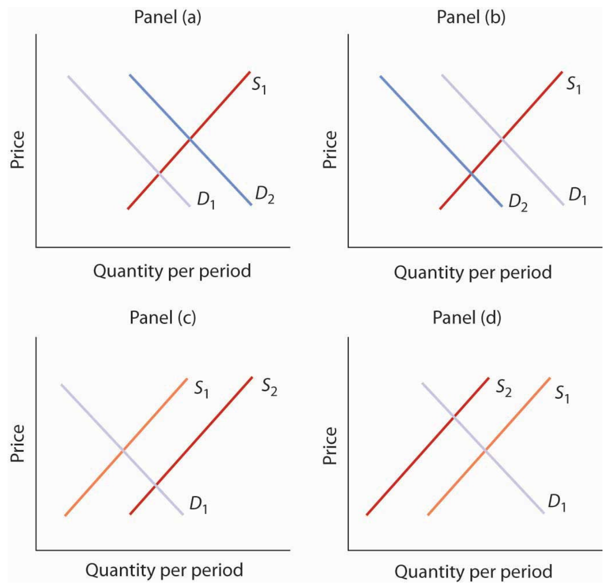
Summary

In this chapter we have examined the model of demand and supply. We found that a demand curve shows the quantity demanded at each price, all other things unchanged. The law of demand asserts that an increase in price reduces the quantity demanded and a decrease in price increases the quantity demanded, all other things unchanged. The supply curve shows the quantity of a good or service that sellers will offer at various prices, all other things unchanged. Supply curves are generally upward sloping: an increase in price generally increases the quantity supplied, all other things unchanged.

The equilibrium price occurs where the demand and supply curves intersect. At this price, the quantity demanded equals the quantity supplied. A price higher than the equilibrium price increases the quantity supplied and reduces the quantity demanded, causing a surplus. A price lower than the equilibrium price increases the quantity demanded and reduces the quantity supplied, causing a shortage. Usually, market surpluses and shortages are short-lived. Changes in demand or supply, caused by changes in the determinants of demand and supply otherwise held constant in the analysis, change the equilibrium price and output. The circular flow model allows us to see how demand and supply in various markets are related to one another.

CONCEPT PROBLEMS

1. What do you think happens to the demand for pizzas during the Super Bowl? Why?
2. Which of the following goods are likely to be classified as normal goods or services? Inferior? Defend your answer.
   a. Beans
   b. Tuxedos
   c. Used cars
   d. Used clothing
   e. Computers
   f. Books reviewed in *The New York Times*
   g. Macaroni and cheese
   h. Calculators
   i. Cigarettes
   j. Caviar
   k. Legal services
3. Which of the following pairs of goods are likely to be classified as substitutes? Complements? Defend your answer.
   a. Peanut butter and jelly
   b. Eggs and ham
   c. Nike brand and Reebok brand sneakers
   d. IBM and Apple Macintosh brand computers
   e. Dress shirts and ties
   f. Airline tickets and hotels
   g. Gasoline and tires
   h. Beer and wine
   i. Faxes and first-class mail
   j. Cereal and milk
   k. Cereal and eggs
4. A study found that lower airfares led some people to substitute flying for driving to their vacation destinations. This reduced the demand for car travel and led to reduced traffic fatalities, since air travel is safer per passenger mile than car travel. Using the logic suggested by that study, suggest how each of the following events would affect the number of highway fatalities in any one year.
   a. An increase in the price of gasoline
   b. A large reduction in rental rates for passenger vans
   c. An increase in airfares
5. Children under age 2 are now allowed to fly free on U.S. airlines; they usually sit in their parents' laps. Some safety advocates have urged that they be required to be strapped in infant seats, which would mean their parents would have to purchase tickets for them. Some economists have argued that such a measure would actually increase infant fatalities. Can you say why?
6. The graphs below show four possible shifts in demand or in supply that could occur in particular markets. Relate each of the events described below to one of them.

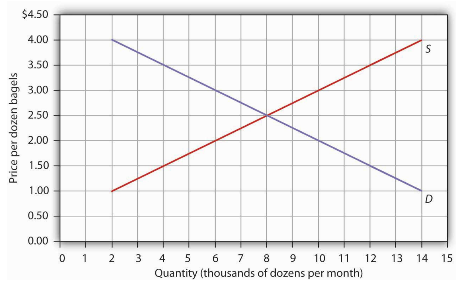


a. How did the heavy rains in South America in 1997 affect the market for coffee?
b. The Surgeon General decides french fries are not bad for your health after all and issues a report endorsing their use. What happens to the market for french fries?
c. How do you think rising incomes affect the market for ski vacations?
d. A new technique is discovered for manufacturing computers that greatly lowers their production cost. What happens to the market for computers?
e. How would a ban on smoking in public affect the market for cigarettes?

7. As low-carb diets increased in popularity, egg prices rose sharply. How might this affect the monks' supply of cookies or private retreats? (See the Case in Point on the Monks of St. Benedict's.)
8. Gasoline prices typically rise during the summer, a time of heavy tourist traffic. A "street talk" feature on a radio station sought tourist reaction to higher gasoline prices. Here was one response: "I don't like 'em [the higher prices] much. I think the gas companies just use any excuse to jack up prices, and they're doing it again now." How does this tourist's perspective differ from that of economists who use the model of demand and supply?
9. The introduction to the chapter argues that preferences for coffee changed in the 1990s and that excessive rain hurt yields from coffee plants. Show and explain the effects of these two circumstances on the coffee market.
10. With preferences for coffee remaining strong in the early part of the century, Vietnam entered the market as a major exporter of coffee. Show and explain the effects of these two circumstances on the coffee market.
11. The study on the economics of obesity discussed in the Case in Point in this chapter on that topic also noted that another factor behind rising obesity is the decline in cigarette smoking as the price of cigarettes has risen. Show and explain the effect of higher cigarette prices on the market for food. What does this finding imply about the relationship between cigarettes and food?
12. In 2004, *The New York Times* reported that India might be losing its outsourcing edge due to rising wages[1] The reporter noted that a recent report "projected that if India continued to produce college graduates at the current rate, demand would exceed supply by 20% in the main outsourcing markets by 2008." Using the terminology you learned in this chapter, explain what he meant to say was happening in the market for Indian workers in outsourcing jobs. In particular, is demand for Indian workers increasing or decreasing? Is the supply of Indian workers increasing or decreasing? Which is shifting faster? How do you know?
13. For more than a century, milk producers have produced skim milk, which contains virtually no fat, along with regular milk, which contains 4% fat. But a century ago, skim milk accounted for only about 1% of total production, and much of it was fed to hogs. Today, skim and other reduced-fat milks make up the bulk of milk sales. What curve shifted, and what factor shifted it?
14. Suppose firms in the economy were to produce fewer goods and services. How do you think this would affect household spending on goods and services? (*Hint:* Use the circular flow model to analyze this question.)

NUMERICAL PROBLEMS

Problems 1–5 are based on the graph below.



1. At a price of $1.50 per dozen, how many bagels are demanded per month?
2. At a price of $1.50 per dozen, how many bagels are supplied per month?
3. At a price of $3.00 per dozen, how many bagels are demanded per month?
4. At a price of $3.00 per dozen, how many bagels are supplied per month?
5. What is the equilibrium price of bagels? What is the equilibrium quantity per month?

Problems 6–9 are based on the model of demand and supply for coffee as shown in Figure 3.17 You can graph the initial demand and supply curves by using the following values, with all quantities in millions of pounds of coffee per month:

Price	Quantity demanded	Quantity supplied
$3	40	10
4	35	15
5	30	20
6	25	25
7	20	30
8	15	35
9	10	40

6. Suppose the quantity demanded rises by 20 million pounds of coffee per month at each price. Draw the initial demand and supply curves based on the values given in the table above. Then draw the new demand curve given by this change, and show the new equilibrium price and quantity.
7. Suppose the quantity demanded falls, relative to the values given in the above table, by 20 million pounds per month at prices between $4 and $6 per pound; at prices between $7 and $9 per pound, the quantity demanded becomes zero. Draw the new demand curve and show the new equilibrium price and quantity.
8. Suppose the quantity supplied rises by 20 million pounds per month at each price, while the quantities demanded retain the values shown in the table above. Draw the new supply curve and show the new equilibrium price and quantity.
9. Suppose the quantity supplied falls, relative to the values given in the table above, by 20 million pounds per month at prices above $5; at a price of $5 or less per pound, the quantity supplied becomes zero. Draw the new supply curve and show the new equilibrium price and quantity.

Problems 10–15 are based on the demand and supply schedules for gasoline below (all quantities are in thousands of gallons per week):

Price per gallon	Quantity demanded	Quantity supplied
$1	8	0
2	7	1
3	6	2
4	5	3
5	4	4
6	3	5
7	2	6
8	1	7

10. Graph the demand and supply curves and show the equilibrium price and quantity.
11. At a price of $3 per gallon, would there be a surplus or shortage of gasoline? How much would the surplus or shortage be? Indicate the surplus or shortage on the graph.
12. At a price of $6 per gallon, would there be a surplus or shortage of gasoline? How much would the surplus or shortage be? Show the surplus or shortage on the graph.
13. Suppose the quantity demanded increased by 2,000 gallons per month at each price. At a price of $3 per gallon, how much would the surplus or shortage be? Graph the demand and supply curves and show the surplus or shortage.
14. Suppose the quantity supplied decreased by 2,000 gallons per month at each price for prices between $4 and $8 per gallon. At prices less than $4 per gallon the quantity supplied becomes zero, while the quantities demanded retain the values shown in the table. At a price of $4 per gallon, how much would the surplus or shortage be? Graph the demand and supply curves and show the surplus or shortage.
15. If the demand curve shifts as in problem 13 and the supply curve shifts as in problem 14, without drawing a graph or consulting the data, can you predict whether equilibrium price increases or decreases? What about equilibrium quantity? Now draw a graph that shows what the new equilibrium price and quantity are.

ENDNOTES

1. Noam Scheiber, “As a Center for Outsourcing, India Could Be Losing Its Edge,” *New York Times*, May 9, 2004, p. BU3.

CHAPTER 4
Applications of Demand and Supply

START UP: A COMPOSER LOGS ON

"Since the age of seven, I knew that I would be a musician. And from age fourteen, I knew that I would be a composer," says Israeli-born Ofer Ben-Amots. What he did not know was that he would use computers to carry out his work. He is now a professor of music at Colorado College, and Dr. Ben-Amots's compositions and operas have been performed in the United States, Europe, and Japan.

For over 15 years, he has used musical notation software to help in composing music. "The output is extremely elegant. Performers enjoy looking at such a clear and clean score. The creation of parts out of a full score is as easy as pressing the <ENTER> key on the keyboard." Changes can easily be inserted into the notation file, which eliminates the need for recopying. In addition, Dr. Ben-Amots uses computers for playback. "I can listen to a relatively accurate 'digital performance' of the score at any given point, with any tempo or instrumentation I choose. The sound quality has improved so much that digital files sound almost identical to real performance." He can also produce CDs on his own and create Podcasts so that anyone in the world can hear his music. He engages in self-publication of scores and self-marketing. "In my case, I get to keep the copyrights on all of my music. This would have been impossible ten to twelve years ago when composers transferred their rights to publishers. Home pages on the World Wide Web allow me to promote my own work." Professor Ben-Amots also changed the way he teaches music composition. New application software, such as GarageBand, has opened the way for anyone interested to try to compose music. Whereas his music composition classes used to have music theory prerequisites, today his classes are open to all.

Dr. Ben-Amots started out in 1989 with a Macintosh SE30 that had 4 megabytes of random access memory (RAM) and an 80-megabyte hard drive. It cost him about $3,000. Today, he uses a Macintosh Powerbook G4 laptop with 1.5 gigabytes of memory, built-in DVD/CD burner, and wireless Internet connections. His new computer cost about $2,000. How personal computers rose so dramatically in power as they fell so steeply in price is just one of the stories about markets we will tell in this chapter, which aims to help you understand how the model of demand and supply applies to the real world.

In the first section of this chapter, we will look at several markets that you are likely to have participated in or be familiar with—the market for personal computers, the markets for crude oil and for gasoline, and the stock market. You probably own or have access to a computer. Each of us was affected by the sharp rise in crude oil and gasoline prices from 2004 to mid-2008. The performance of the stock market is always a major news item and may affect you personally, if not now, then in the future. The concepts of demand and supply go a long way in explaining the behavior of equilibrium prices and quantities in all of these markets. The purpose of this section is to allow you to practice using the model of demand and supply and get you to start thinking about the myriad ways the model of demand and supply can be applied.

In the second part of the chapter we will look at markets in which the government has historically played a large role in regulating prices. By legislating maximum or minimum prices, the government has kept the prices of certain goods below or above equilibrium. We will look at the arguments for direct government intervention in controlling prices as well as the consequences of such policies. As we shall see, preventing the price of a good from finding its own equilibrium often has consequences that may be at odds with the intentions of the policy makers who put the regulations in place.

In the third section of the chapter we will look at the market for health care. This market is interesting because how well (or poorly) it works can be a matter of life and death and because it has special characteristics. In particular, markets in which participants do not pay for goods directly, but rather pay insurers who then pay the suppliers of the goods, operate somewhat differently from those in which participants pay directly for their purchases. This extension of demand and supply analysis reveals much about how such markets operate.

1. PUTTING DEMAND AND SUPPLY TO WORK

LEARNING OBJECTIVES

1. **Learn how to apply the model of demand and supply to explaining the behavior of equilibrium prices and quantities in a variety of markets.**
2. **Explain how technological change can be represented using the model of demand and supply.**
3. **Explain how the model of demand and supply can be used to explain changes in prices of shares of stock.**

A shift in either demand or supply, or in both, leads to a change in equilibrium price and equilibrium quantity. We begin this chapter by examining markets in which prices adjust quickly to changes in demand or supply: the market for personal computers, the markets for crude oil and gasoline, and the stock market. These markets are thus direct applications of the model of demand and supply.

1.1 The Personal Computer Market

In the 1960s, to speak of computers was to speak of IBM, the dominant maker of large mainframe computers used by business and government agencies. Then between 1976, when Apple Computer introduced its first desktop computer, and 1981, when IBM produced its first personal computers (PCs), the old world was turned upside down. In 1984, just 8.2% of U.S. households owned a personal computer. By 2007, Google estimates that 78% did. The tools of demand and supply tell the story from an economic perspective.

Technological change has been breathtakingly swift in the computer industry. Because personal computers have changed so dramatically in performance and in the range of the functions they perform, we shall speak of "quality-adjusted" personal computers. The price per unit of quality-adjusted desktop computers fell by about half every 50 months during the period 1976–1989. In the first half of the 1990s, those prices fell by half every 28 months. In the second half of the 1990s, the "halving time" fell to every 24 months.[1]

Consider another indicator of the phenomenal change in computers. Between 1993 and 1998, the Bureau of Labor Statistics estimates that central processing unit (CPU) speed rose 1,263%, system memory increased 1,500%, hard drive capacity soared by 3,700%, and monitor size went up 13%. It seems safe to say that the dizzying pace of change recorded in the 1990s has increased in this century. A "computer" today is not the same good as a "computer" even five years ago. To make them comparable, we must adjust for these changes in quality.

Initially, most personal computers were manufactured by Apple or Compaq; both companies were very profitable. The potential for profits attracted IBM and other firms to the industry. Unlike large mainframe computers, personal computer clones turned out to be fairly easy things to manufacture. As shown in Table 4.1, the top five personal computer manufacturers produced only 48% of the personal computers sold in the world in 2005, and the largest manufacturer, Dell, sold only about 19% of the total in that year. This is a far cry from the more than 90% of the mainframe computer market that IBM once held. The market has become far more competitive.

TABLE 4.1 Personal Computer Shipments, Market Percentage Shares by Vendors, World and United States

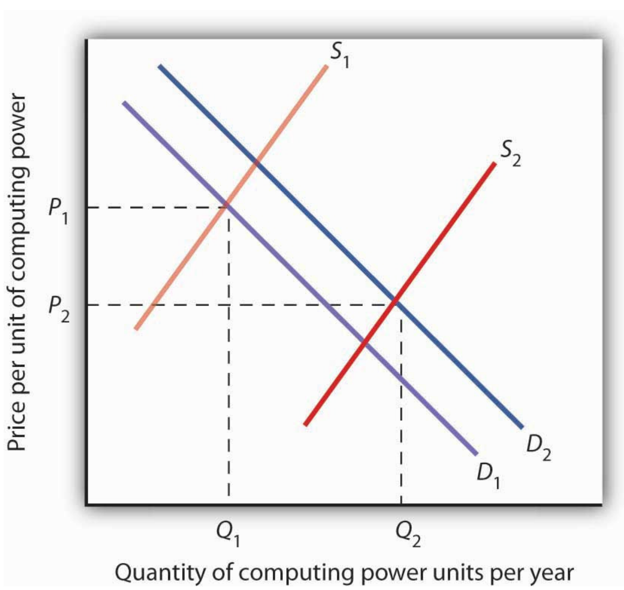
Company	% of World Shipments	Company	% of U.S. Shipments
Dell	18.9	Dell	34
Hewlett-Packard	15.4	Hewlett-Packard	18.2
IBM	5.1	Gateway	5.7
Fujitsu Seimens	4.6	IBM	4.3
Acer	4	Apple	3.9
Others	52	Others	34
Total	**100.0**	**Total**	**100.0**

Source: IDC—Press Release 15 Apr 2005 "PC Market Approaches 11% Growth as International Demand Remains Strong, According to IDC" (http://www.idc.com/getdoc.jsp?containerId=pr2005_04_14_17070722) (Totals may not add due to rounding)

Figure 4.1 illustrates the changes that have occurred in the computer market. The horizontal axis shows the quantity of quality-adjusted personal computers. Thus, the quantity axis can be thought of as a unit of computing power. Similarly, the price axis shows the price per unit of computing power. The rapid increase in the number of firms, together with dramatic technological improvements, led to an increase in supply, shifting the supply curve in Figure 4.1 to the right from S_1 to S_2.

FIGURE 4.1 The Personal Computer Market

The supply curve for quality-adjusted personal computers has shifted markedly to the right, reducing the equilibrium price from P_1 to P_2 and increasing the equilibrium quantity from Q_1 to Q_2 in 2005.



Demand also shifted to the right from D_1 to D_2, as incomes rose and new uses for computers, from e-mail and social networking to Voice over Internet Protocol (VoIP) and Radio Frequency ID (RFID) tags (which allow wireless tracking of commercial shipments via desktop computers), altered the preferences of consumer and business users. Because we observe a fall in equilibrium price and an increase in equilibrium quantity, we conclude that the rightward shift in supply has outweighed the rightward shift in demand. The power of market forces has profoundly affected the way we live and work.

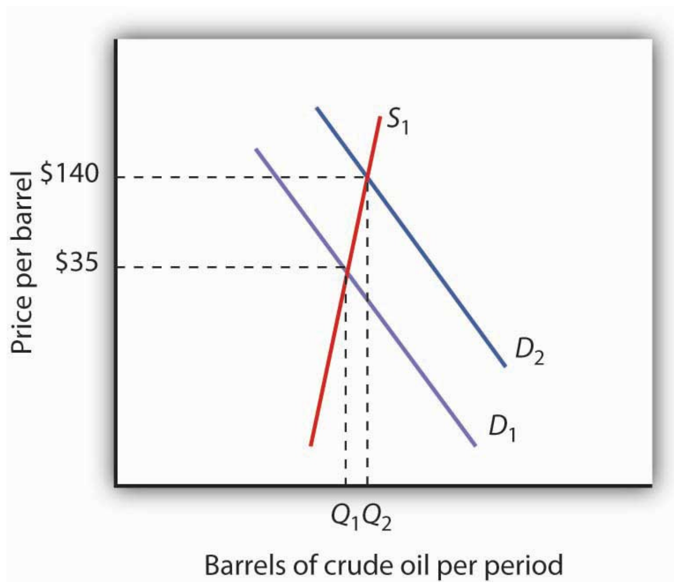
1.2 The Markets for Crude Oil and for Gasoline

The market for crude oil took a radical turn in 1973. The price per barrel of crude oil quadrupled in 1973 and 1974. Price remained high until the early 1980s but then fell back drastically and remained low for about two decades. In 2004, the price of oil began to move upward and by 2008 had reached $147 per barrel.

What caused the dramatic increase in gasoline and oil prices in 2008? It appeared to be increasing worldwide demand outpacing producers' ability—or willingness—to increase production much. This increase in demand is illustrated in Figure 4.2.

FIGURE 4.2 The Increasing Demand for Crude Oil

The price of oil was $35 per barrel at the beginning of 2004, as determined by the intersection of world demand, D_1, and world supply, S_1. Increasing world demand, prompted largely by increasing demand from China as well as from other countries, shifted world demand to D_2, pushing the price as high as $140 per barrel by the middle of 2008.

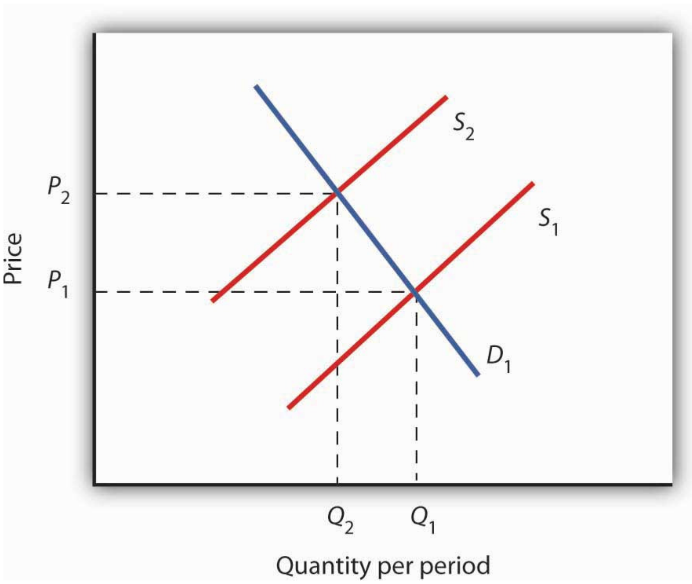


Higher oil prices also increase the cost of producing virtually every good or service, as at a minimum, the production of most goods requires transportation. These costs inevitably translate into higher prices for nearly all goods and services. Supply curves of the goods and services thus affected shift to the left, putting downward pressure on output and upward pressure on prices.

Graphically, the impact of higher gasoline prices on businesses that use gasoline is illustrated in Figure 4.3. Because higher gasoline prices increase the cost of doing business, they shift the supply curves for nearly all businesses to the left, putting upward pressure on prices and downward pressure on output. In the case shown here, the supply curve in a typical industry shifts from S_1 to S_2. This increases the equilibrium price from P_1 to P_2 and reduces the equilibrium quantity from Q_1 to Q_2.

FIGURE 4.3 The Impact of Higher Gasoline Prices

Higher gasoline prices increase the cost of producing virtually every good or service. In the case shown here, the supply curve in a typical industry shifts from S_1 to S_2. This increases the equilibrium price from P_1 to P_2 and reduces equilibrium quantity from Q_1 to Q_2.



Then, as the world economy slowed dramatically in the second half of 2008, the demand curve for oil shifted back to the left. By November 2008, the price per barrel had dropped back to below $60 per barrel. As gas prices also subsided, so did the threat of higher prices in other industries.

1.3 The Stock Market

The circular flow model suggests that capital, like other factors of production, is supplied by households to firms. Firms, in turn, pay income to those households for the use of their capital. Generally speaking, however, capital is actually owned by firms themselves. General Motors owns its assembly plants, and Wal-Mart owns its stores; these firms therefore own their capital. But firms, in turn, are owned by people—and those people, of course, live in households. It is through their ownership of firms that households own capital.

A firm may be owned by one individual (a **sole proprietorship**), by several individuals (a **partnership**), or by shareholders who own stock in the firm (a **corporation**). Although most firms in the United States are sole proprietorships or partnerships, the bulk of the nation's total output (about 90%) is produced by corporations. Corporations also own most of the capital (machines, plants, buildings, and the like).

This section describes how the prices of shares of **corporate stock**, shares in the ownership of a corporation, are determined by the interaction of demand and supply. Ultimately, the same forces that determine the value of a firm's stock determine the value of a sole proprietorship or partnership.

When a corporation needs funds to increase its capital or for other reasons, one means at its disposal is to issue new stock in the corporation. (Other means include borrowing funds or using past profits.) Once the new shares have been sold in what is called an initial public offering (IPO), the corporation receives no further funding as shares of its stock are bought and sold on the secondary market. The secondary market is the market for stocks that have been issued in the past, and the daily news reports about stock prices almost always refer to activity in the secondary market. Generally, the corporations whose shares are traded are not involved in these transactions.

The **stock market** is the set of institutions in which shares of stock are bought and sold. The New York Stock Exchange (NYSE) is one such institution. There are many others all over the world, such as the DAX in Germany and the Bolsa in Mexico. To buy or sell a share of stock, one places an order with a stockbroker who relays the order to one of the traders at the NYSE or at some other exchange.

The process through which shares of stock are bought and sold can seem chaotic. At many exchanges, traders with orders from customers who want to buy stock shout out the prices those customers are willing to pay. Traders with orders from customers who want to sell shout out offers of prices at which their customers are willing to sell. Some exchanges use electronic trading, but the principle is the same: if the price someone is willing to pay matches the price at which someone else is willing to sell,

sole proprietorship

A firm owned by one individual.

partnership

A firm owned by several individuals.

corporation

A firm owned by shareholders who own stock in the firm.

corporate stock

Shares in the ownership of a corporation.

stock market

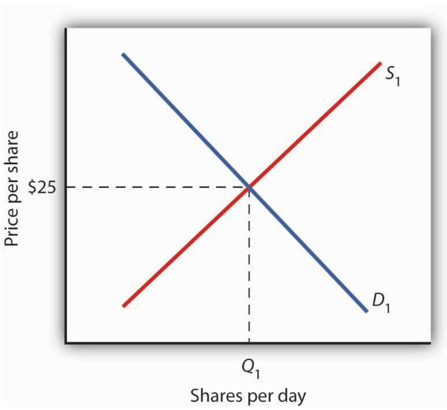
The set of institutions in which shares of stock are bought and sold.

the trade is made. The most recent price at which a stock has traded is reported almost instantaneously throughout the world.

Figure 4.4 applies the model of demand and supply to the determination of stock prices. Suppose the demand curve for shares in Intel Corporation is given by D_1 and the supply by S_1. (Even though the total number of shares outstanding is fixed at any point in time, the supply curve is not vertical. Rather, the supply curve is upward sloping because it represents how many shares current owners are prepared to sell at each price, and that number will be greater at higher prices.) Suppose that these curves intersect at a price of $25, at which Q_1 shares are traded each day. If the price were higher, more shares would be offered for sale than would be demanded, and the price would quickly fall. If the price were lower, more shares would be demanded than would be supplied, and the price would quickly rise. In general, we can expect the prices of shares of stock to move quickly to their equilibrium levels.

FIGURE 4.4 Demand and Supply in the Stock Market

The equilibrium price of stock shares in Intel Corporation is initially $25, determined by the intersection of demand and supply curves D_1 and S_1, at which Q_1 million shares are traded each day.



The intersection of the demand and supply curves for shares of stock in a particular company determines the equilibrium price for a share of stock. But what determines the demand and supply for shares of a company's stock?

The owner of a share of a company's stock owns a share of the company, and, hence, a share of its profits; typically, a corporation will retain and reinvest some of its profits to increase its future profitability. The profits kept by a company are called **retained earnings**. Profits distributed to shareholders are called **dividends**. Because a share of stock gives its owner a claim on part of a company's future profits, it follows that the expected level of future profits plays a role in determining the value of its stock.

Of course, those future profits cannot be known with certainty; investors can only predict what they might be, based on information about future demand for the company's products, future costs of production, information about the soundness of a company's management, and so on. Stock prices in the real world thus reflect estimates of a company's profits projected into the future.

The downward slope of the demand curve suggests that at lower prices for the stock, more people calculate that the firm's future earnings will justify the stock's purchase. The upward slope of the supply curve tells us that as the price of the stock rises, more people conclude that the firm's future earnings do not justify holding the stock and therefore offer to sell it. At the equilibrium price, the number of shares supplied by people who think holding the stock no longer makes sense just balances the number of shares demanded by people who think it does.

What factors, then, cause the demand or supply curves for shares of stocks to shift? The most important factor is a change in the expectations of a company's future profits. Suppose Intel announces a new generation of computer chips that will lead to faster computers with larger memories. Current owners of Intel stock would adjust upward their estimates of what the value of a share of Intel stock should be. At the old equilibrium price of $25 fewer owners of Intel stock would be willing to sell. Since this would be true at every possible share price, the supply curve for Intel stock would shift to the left, as shown in Figure 4.5. Just as the expectation that a company will be more profitable shifts the supply curve for its stock to the left, that same change in expectations will cause more people to want to purchase the stock, shifting the demand curve to the right. In Figure 4.5, we see the supply curve shifting to the left, from S_1 to S_2, while the demand curve shifts to the right, from D_1 to D_2.

retained earnings

Profits kept by a company.

dividends

Profits distributed to shareholders.

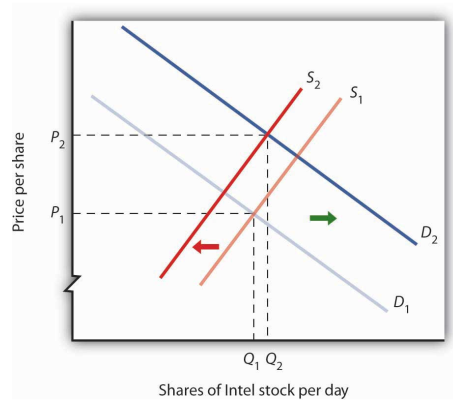
Other factors may alter the price of an individual corporation's share of stock or the level of stock prices in general. For example, demographic change and rising incomes have affected the demand for stocks in recent years. For example, with a large proportion of the U.S. population nearing retirement age and beginning to think about and plan for their lives during retirement, the demand for stocks has risen.

Information on the economy as a whole is also likely to affect stock prices. If the economy overall is doing well and people expect that to continue, they may become more optimistic about how profitable companies will be in general, and thus the prices of stocks will rise. Conversely, expectations of a sluggish economy, as happened in the fall of 2008, could cause stock prices in general to fall.

The stock market is bombarded with new information every minute of every day. Firms announce their profits of the previous quarter. They announce that they plan to move into a new product line or sell their goods in another country. We learn that the price of Company A's good, which is a substitute for one sold by Company B, has risen. We learn that countries sign trade agreements, launch wars, or make peace. All of this information may affect stock prices because any information can affect how buyers and sellers value companies.

FIGURE 4.5 A Change in Expectations Affects the Price of Corporate Stock

If financial investors decide that a company is likely to be more profitable, then the supply of the stock shifts to the left (in this case, from S_1 to S_2), and the demand for the stock shifts to the right (in this case, from D_1 to D_2), resulting in an increase in price from P_1 to P_2.



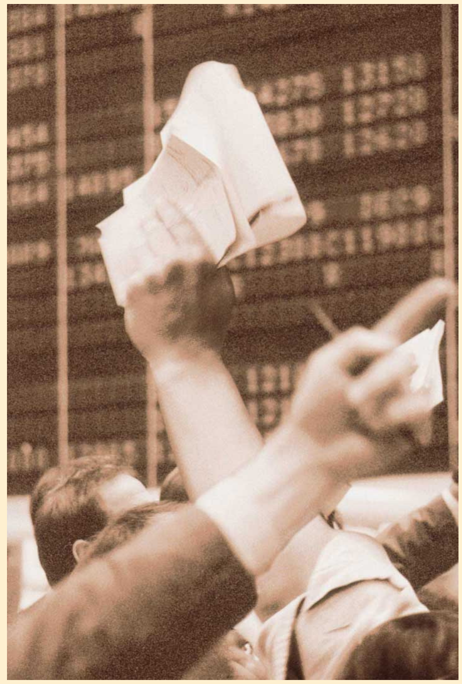
KEY TAKEAWAYS

- Technological change, which has caused the supply curve for computing power to shift to the right, is the main reason for the rapid increase in equilibrium quantity and decrease in equilibrium price of personal computers.
- The increase in crude oil and gasoline prices in 2008 was driven primarily by increased demand for crude oil, an increase that was created by economic growth throughout the world. Crude oil and gas prices fell markedly as world economic growth subsided later in the year.
- Higher gasoline prices increased the cost of producing virtually every good and service, shifting supply curves for most goods and services to the left. This tended to push prices up and output down.
- Demand and supply determine prices of shares of corporate stock. The equilibrium price of a share of stock strikes a balance between those who think the stock is worth more and those who think it is worth less than the current price.
- If a company's profits are expected to increase, the demand curve for its stock shifts to the right and the supply curve shifts to the left, causing equilibrium price to rise. The opposite would occur if a company's profits were expected to decrease.
- Other factors that influence the price of corporate stock include demographic and income changes and the overall health of the economy.

TRY IT!

Suppose an airline announces that its earnings this year are lower than expected due to reduced ticket sales. The airline spokesperson gives no information on how the company plans to turn things around. Use the model of demand and supply to show and explain what is likely to happen to the price of the airline's stock.

Case in Point: 9/11 and the Stock Market



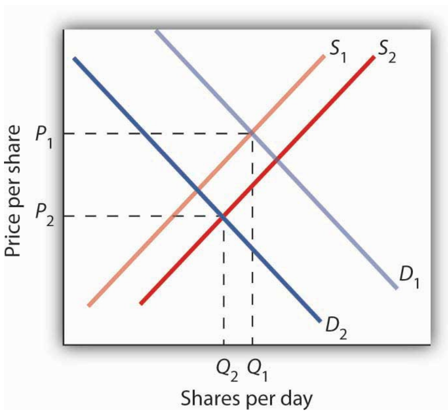
The hijacking of four airplanes and the steering of them into buildings is perhaps the only disaster that has become universally known by its date: September 11, 2001—hence, 9/11. "9/11" will remain etched in our collective memory for a great many generations.

Disasters such as 9/11 represent the kind of complete surprises that dramatically affect stock prices, if only temporarily. The New York Stock Exchange was closed on the day of the attack and remained closed for six days. On the day the market opened, the Dow Jones Industrial Average (the "DOW", a widely used gauge of stock prices) fell nearly 685 points to 8,920. It was one of the biggest one-day decline in U.S. history.

Why did the attacks on September 11, 2001, have such a dramatic short-term impact on the stock market? The attacks of 9/11 plunged the United States and much of the rest of the world into a very frightening war against terrorism. The realization that terrorists could strike anytime and in any place sapped consumer and business confidence alike and affected both the demand and supply of most stocks. The attacks on 9/11 provoked fear and uncertainty—two things that are certain to bring stock prices down, at least until other events and more information cause expectations to change again in this very responsive market.

ANSWER TO TRY IT! PROBLEM

The information given in the problem suggests that the airline's profits are likely to fall below expectations. Current owners of the airline's stock and potential buyers of the stock would adjust downward their estimates of what the value of the corporation's stock should be. As a result the supply curve for the stock would increase, shifting it to the right, while the demand curve for the stock would decrease, shifting it to the left. As a result, equilibrium price of the stock falls from P_1 to P_2. What happens to equilibrium quantity depends on the extent to which each curve shifts. In the diagram, equilibrium quantity is shown to decrease from Q_1 to Q_2.



2. GOVERNMENT INTERVENTION IN MARKET PRICES: PRICE FLOORS AND PRICE CEILINGS

LEARNING OBJECTIVES

1. **Use the model of demand and supply to explain what happens when the government imposes price floors or price ceilings.**
2. **Discuss the reasons why governments sometimes choose to control prices and the consequences of price control policies.**

So far in this chapter and in the previous chapter, we have learned that markets tend to move toward their equilibrium prices and quantities. Surpluses and shortages of goods are short-lived as prices adjust to equate quantity demanded with quantity supplied.

In some markets, however, governments have been called on by groups of citizens to intervene to keep prices of certain items higher or lower than what would result from the market finding its own equilibrium price. In this section we will examine agricultural markets and apartment rental markets—two markets that have often been subject to price controls. Through these examples, we will identify the effects of controlling prices. In each case, we will look at reasons why governments have chosen to control prices in these markets and the consequences of these policies.

2.1 Agricultural Price Floors

Governments often seek to assist farmers by setting price floors in agricultural markets. A minimum allowable price set above the equilibrium price is a **price floor**. With a price floor, the government forbids a price below the minimum. (Notice that, if the price floor were for whatever reason set below the equilibrium price, it would be irrelevant to the determination of the price in the market since nothing would prohibit the price from rising to equilibrium.) A price floor that is set above the equilibrium price creates a surplus.

price floor

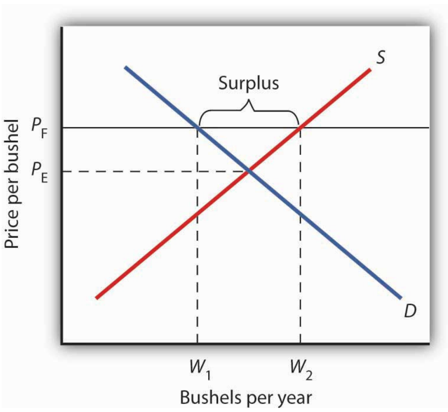
A minimum allowable price set above the equilibrium price.

Figure 4.8 shows the market for wheat. Suppose the government sets the price of wheat at P_F. Notice that P_F is above the equilibrium price of P_E. At P_F, we read over to the demand curve to find that the quantity of wheat that buyers will be willing and able to purchase is W_1 bushels. Reading over to the supply curve, we find that sellers will offer W_2 bushels of wheat at the price floor of P_F. Because

P_F is above the equilibrium price, there is a surplus of wheat equal to (W_2 – W_1) bushels. The surplus persists because the government does not allow the price to fall.

FIGURE 4.8 Price Floors in Wheat Markets

A price floor for wheat creates a surplus of wheat equal to (W_2 - W_1) bushels.



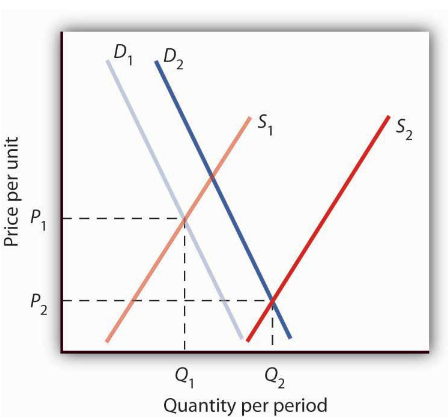
Why have many governments around the world set price floors in agricultural markets? Farming has changed dramatically over the past two centuries. Technological improvements in the form of new equipment, fertilizers, pesticides, and new varieties of crops have led to dramatic increases in crop output per acre. Worldwide production capacity has expanded markedly. As we have learned, technological improvements cause the supply curve to shift to the right, reducing the price of food. While such price reductions have been celebrated in computer markets, farmers have successfully lobbied for government programs aimed at keeping their prices from falling.

While the supply curve for agricultural goods has shifted to the right, the demand has increased with rising population and with rising income. But as incomes rise, people spend a smaller and smaller fraction of their incomes on food. While the demand for food has increased, that increase has not been nearly as great as the increase in supply. Figure 4.9 shows that the supply curve has shifted much farther to the right, from S_1 to S_2, than the demand curve has, from D_1 to D_2. As a result, equilibrium quantity has risen dramatically, from Q_1 to Q_2, and equilibrium price has fallen, from P_1 to P_2.

On top of this long-term historical trend in agriculture, agricultural prices are subject to wide swings over shorter periods. Droughts or freezes can sharply reduce supplies of particular crops, causing sudden increases in prices. Demand for agricultural goods of one country can suddenly dry up if the government of another country imposes trade restrictions against its products, and prices can fall. Such dramatic shifts in prices and quantities make incomes of farmers unstable.

FIGURE 4.9 Supply and Demand Shifts for Agricultural Products

A relatively large increase in the supply of agricultural products, accompanied by a relatively small increase in demand, has reduced the price received by farmers and increased the quantity of agricultural goods.



The Great Depression of the 1930s led to a major federal role in agriculture. The Depression affected the entire economy, but it hit farmers particularly hard. Prices received by farmers plunged nearly two-thirds from 1930 to 1933. Many farmers had a tough time keeping up mortgage payments. By 1932, more than half of all farm loans were in default.

Farm legislation passed during the Great Depression has been modified many times, but the federal government has continued its direct involvement in agricultural markets. This has meant a variety of government programs that guarantee a minimum price for some types of agricultural products. These programs have been accompanied by government purchases of any surplus, by requirements to restrict acreage in order to limit those surpluses, by crop or production restrictions, and the like.

To see how such policies work, look back at Figure 4.8. At P_F, W_2 bushels of wheat will be supplied. With that much wheat on the market, there is market pressure on the price of wheat to fall. To prevent price from falling, the government buys the surplus of (W_2 - W_1) bushels of wheat, so that only W_1 bushels are actually available to private consumers for purchase on the market. The government can store the surpluses or find special uses for them. For example, surpluses generated in the United States have been shipped to developing countries as grants-in-aid or distributed to local school lunch programs. As a variation on this program, the government can require farmers who want to participate in the price support program to reduce acreage in order to limit the size of the surpluses.

After 1973, the government stopped buying the surpluses (with some exceptions) and simply guaranteed farmers a "target price." If the average market price for a crop fell below the crop's target price, the government paid the difference. If, for example, a crop had a market price of \$3 per unit and a target price of \$4 per unit, the government would give farmers a payment of \$1 for each unit sold. Farmers would thus receive the market price of \$3 plus a government payment of \$1 per unit. For farmers to receive these payments, they had to agree to remove acres from production and to comply with certain conservation provisions. These restrictions sought to reduce the size of the surplus generated by the target price, which acted as a kind of price floor.

What are the effects of such farm support programs? The intention is to boost and stabilize farm incomes. But, with price floors, consumers pay more for food than they would otherwise, and governments spend heavily to finance the programs. With the target price approach, consumers pay less, but government financing of the program continues. U.S. federal spending for agriculture averaged well over \$22 billion per year between 2003 and 2007, roughly \$70 per person.

Help to farmers has sometimes been justified on the grounds that it boosts incomes of "small" farmers. However, since farm aid has generally been allotted on the basis of how much farms produce rather than on a per-farm basis, most federal farm support has gone to the largest farms. If the goal is to eliminate poverty among farmers, farm aid could be redesigned to supplement the incomes of small or poor farmers rather than to undermine the functioning of agricultural markets.

In 1996, the U.S. Congress passed the Federal Agriculture Improvement and Reform Act of 1996, or FAIR. The thrust of the new legislation was to do away with the various programs of price support for most crops and hence provide incentives for farmers to respond to market price signals. To protect farmers through a transition period, the act provided for continued payments that were scheduled to decline over a seven-year period. However, with prices for many crops falling in 1998, the U.S. Congress passed an emergency aid package that increased payments to farmers. In 2008, as farm prices reached record highs, Congress passed a farm bill that increased subsidy payments to $40 billion. It did, however, for the first time limit payments to the wealthiest farmers. Individual farmers whose farm incomes exceed $750,000 (or $1.5 million for couples) would be ineligible for some subsidy programs.

2.2 Rental Price Ceilings

The purpose of rent control is to make rental units cheaper for tenants than they would otherwise be. Unlike agricultural price controls, rent control in the United States has been largely a local phenomenon, although there were national rent controls in effect during World War II. Currently, about 200 cities and counties have some type of rent control provisions, and about 10% of rental units in the United States are now subject to price controls. New York City's rent control program, which began in 1943, is among the oldest in the country. Many other cities in the United States adopted some form of rent control in the 1970s. Rent controls have been pervasive in Europe since World War I, and many large cities in poorer countries have also adopted rent controls.

Rent controls in different cities differ in terms of their flexibility. Some cities allow rent increases for specified reasons, such as to make improvements in apartments or to allow rents to keep pace with price increases elsewhere in the economy. Often, rental housing constructed after the imposition of the rent control ordinances is exempted. Apartments that are vacated may also be decontrolled. For simplicity, the model presented here assumes that apartment rents are controlled at a price that does not change.

Figure 4.10 shows the market for rental apartments. Notice that the demand and supply curves are drawn to look like all the other demand and supply curves you have encountered so far in this text: the demand curve is downward-sloping and the supply curve is upward-sloping.

The demand curve shows that a higher price (rent) reduces the quantity of apartments demanded. For example, with higher rents, more young people will choose to live at home with their parents. With lower rents, more will choose to live in apartments. Higher rents may encourage more apartment sharing; lower rents would induce more people to live alone.

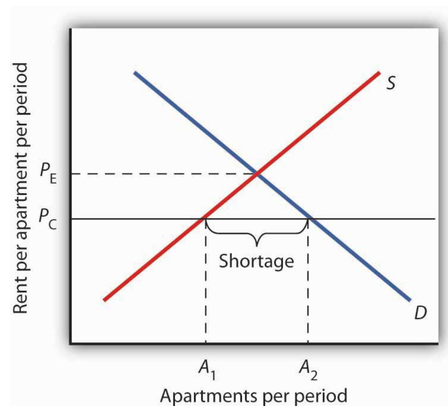
The supply curve is drawn to show that as rent increases, property owners will be encouraged to offer more apartments to rent. Even though an aerial photograph of a city would show apartments to be fixed at a point in time, owners of those properties will decide how many to rent depending on the amount of rent they anticipate. Higher rents may also induce some homeowners to rent out apartment space. In addition, renting out apartments implies a certain level of service to renters, so that low rents may lead some property owners to keep some apartments vacant.

Rent control is an example of a **price ceiling**, a maximum allowable price. With a price ceiling, the government forbids a price above the maximum. A price ceiling that is set below the equilibrium price creates a shortage that will persist.

Suppose the government sets the price of an apartment at P_C in Figure 4.10. Notice that P_C is below the equilibrium price of P_E. At P_C, we read over to the supply curve to find that sellers are willing to offer A_1 apartments. Reading over to the demand curve, we find that consumers would like to rent A_2 apartments at the price ceiling of P_C. Because P_C is below the equilibrium price, there is a shortage of apartments equal to (A_2 - A_1). (Notice that if the price ceiling were set above the equilibrium price it would have no effect on the market since the law would not prohibit the price from settling at an equilibrium price that is lower than the price ceiling.)

FIGURE 4.10 Effect of a Price Ceiling on the Market for Apartments

A price ceiling on apartment rents that is set below the equilibrium rent creates a shortage of apartments equal to ($A_2 - A_1$) apartments.

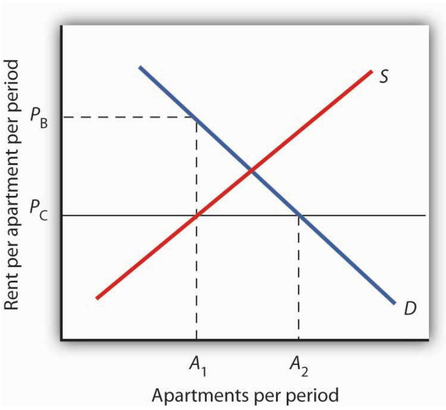


price ceiling

A maximum allowable price.

FIGURE 4.11 The Unintended Consequences of Rent Control

Controlling apartment rents at P_C creates a shortage of ($A_2 - A_1$) apartments. For A_1 apartments, consumers are willing and able to pay P_B, which leads to various "backdoor" payments to apartment owners.



If rent control creates a shortage of apartments, why do some citizens nonetheless clamor for rent control and why do governments often give in to the demands? The reason generally given for rent control is to keep apartments affordable for low- and middle-income tenants.

But the reduced quantity of apartments supplied must be rationed in some way, since, at the price ceiling, the quantity demanded would exceed the quantity supplied. Current occupants may be reluctant to leave their dwellings because finding other apartments will be difficult. As apartments do become available, there will be a line of potential renters waiting to fill them, any of whom is willing to pay the controlled price of P_C or more. In fact, reading up to the demand curve in Figure 4.11 from A_1 apartments, the quantity available at P_C, you can see that for A_1 apartments, there are potential renters willing and able to pay P_B. This often leads to various "backdoor" payments to apartment owners, such as large security deposits, payments for things renters may not want (such as furniture), so-called "key" payments ("The monthly rent is \$500 and the key price is \$3,000"), or simple bribes.

In the end, rent controls and other price ceilings often end up hurting some of the people they are intended to help. Many people will have trouble finding apartments to rent. Ironically, some of those who do find apartments may actually end up paying more than they would have paid in the absence of rent control. And many of the people that the rent controls do help (primarily current occupants, regardless of their income, and those lucky enough to find apartments) are not those they are intended to help (the poor). There are also costs in government administration and enforcement.

Because New York City has the longest history of rent controls of any city in the United States, its program has been widely studied. There is general agreement that the rent control program has reduced tenant mobility, led to a substantial gap between rents on controlled and uncontrolled units, and favored long-term residents at the expense of newcomers to the city.[2] These distortions have grown over time, another frequent consequence of price controls.

A more direct means of helping poor tenants, one that would avoid interfering with the functioning of the market, would be to subsidize their incomes. As with price floors, interfering with the market mechanism may solve one problem, but it creates many others at the same time.

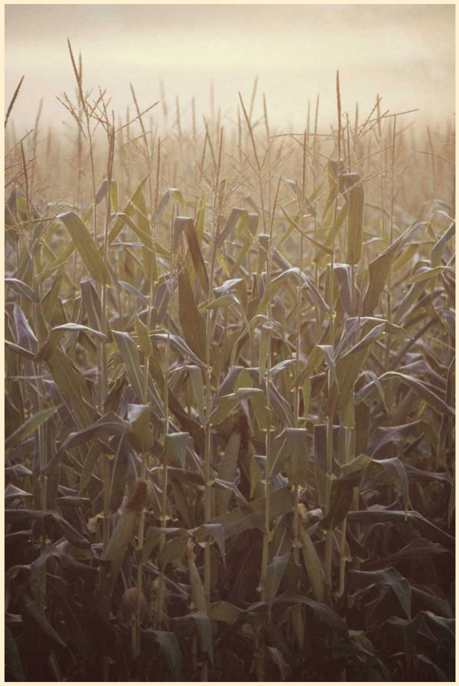
KEY TAKEAWAYS

- Price floors create surpluses by fixing the price above the equilibrium price. At the price set by the floor, the quantity supplied exceeds the quantity demanded.
- In agriculture, price floors have created persistent surpluses of a wide range of agricultural commodities. Governments typically purchase the amount of the surplus or impose production restrictions in an attempt to reduce the surplus.
- Price ceilings create shortages by setting the price below the equilibrium. At the ceiling price, the quantity demanded exceeds the quantity supplied.
- Rent controls are an example of a price ceiling, and thus they create shortages of rental housing.
- It is sometimes the case that rent controls create "backdoor" arrangements, ranging from requirements that tenants rent items that they do not want to outright bribes, that result in rents higher than would exist in the absence of the ceiling.

TRY IT!

A minimum wage law is another example of a price floor. Draw demand and supply curves for unskilled labor. The horizontal axis will show the quantity of unskilled labor per period and the vertical axis will show the hourly wage rate for unskilled workers, which is the price of unskilled labor. Show and explain the effect of a minimum wage that is above the equilibrium wage.

Case in Point: Corn: It Is Not Just Food Any More



Government support for corn dates back to the Agricultural Act of 1938 and, in one form or another, has been part of agricultural legislation ever since. Types of supports have ranged from government purchases of surpluses to target pricing, land set asides, and loan guarantees. According to one estimate, the U.S. government spent nearly $42 billion to support corn between 1995 and 2004.

Then, during the period of rising oil prices of the late 1970s and mounting concerns about dependence on foreign oil from volatile regions in the world, support for corn, not as a food, but rather as an input into the production of ethanol—an alternative to oil-based fuel—began. Ethanol tax credits were part of the Energy Act of 1978. Since 1980, a tariff of 50¢ per gallon against imported ethanol, even higher today, has served to protect domestic corn-based ethanol from imported ethanol, in particular from sugar-cane-based ethanol from Brazil.

The Energy Policy Act of 2005 was another milestone in ethanol legislation. Through loan guarantees, support for research and development, and tax credits, it mandated that 4 billion gallons of ethanol be used by 2006 and 7.5 billion gallons by 2012. Ethanol production had already reached 6.5 billion gallons by 2007, so new legislation in 2007 upped the ante to 15 billion gallons by 2015.

Beyond the increased amount the government is spending to support corn and corn-based ethanol, criticism of the policy has three major prongs:

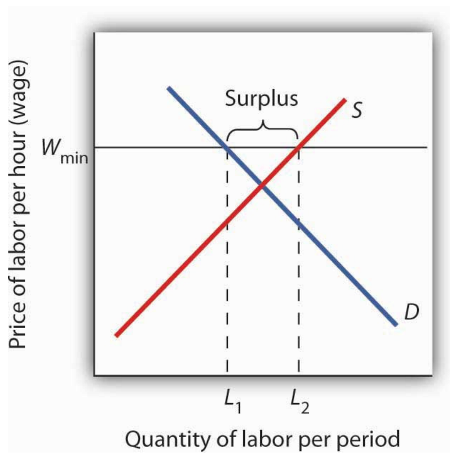
1. Corn-based ethanol does little to reduce U.S. dependence on foreign oil because the energy required to produce a gallon of corn-based ethanol is quite high. A 2006 National Academy of Sciences paper estimated that one gallon of ethanol is needed to bring 1.25 gallons of it to market. Other studies show an even less favorable ratio.
2. Biofuels, such as corn-based ethanol, are having detrimental effects on the environment, with increased deforestation, stemming from more land being used to grow fuel inputs, contributing to global warming.
3. The diversion of corn and other crops from food to fuel is contributing to rising food prices and an increase in world hunger. C. Ford Runge and Benjamin Senauer wrote in *Foreign Affairs* that even small increases in prices of food staples have severe consequences on the very poor of the world, and "Filling the 25-gallon tank of an SUV with pure ethanol requires over 450 pounds of corn—which contains enough calories to feed one person for a year."

Some of these criticisms may be contested as exaggerated: Will the ratio of energy-in to energy-out improve as new technologies emerge for producing ethanol? Did not other factors, such as weather and rising food demand worldwide, contribute to higher grain prices? Nonetheless, it is clear that corn-based ethanol is no free lunch. It is also clear that the end of government support for corn is nowhere to be seen.

Sources: Alexei Barrionuevo, "Mountains of Corn and a Sea of Farm Subsidies," New York Times, November 9, 2005, online version; David Freddoso, "Children of the Corn," National Review Online, May 6, 2008; C. Ford Runge and Benjamin Senauer, "How Biofuels Could Starve the Poor," Foreign Affairs, May/June 2007, online version; Michael Grunwald, "The Clean Energy Scam," Time 171:14 (April 7, 2008): 40–45.

ANSWER TO TRY IT! PROBLEM

A minimum wage (W_{min}) that is set above the equilibrium wage would create a surplus of unskilled labor equal to (L_2 - L_1). That is, L_2 units of unskilled labor are offered at the minimum wage, but companies only want to use L_1 units at that wage. Because unskilled workers are a substitute for a skilled workers, forcing the price of unskilled workers higher would increase the demand for skilled labor and thus increase their wages.



3. THE MARKET FOR HEALTH-CARE SERVICES

LEARNING OBJECTIVE

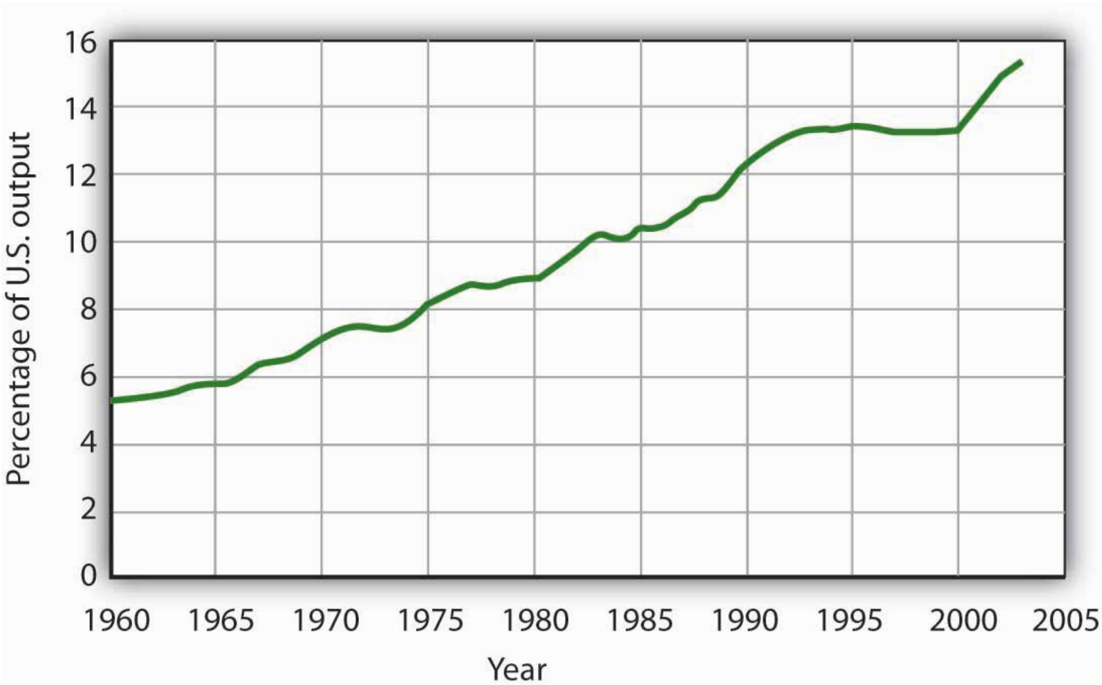
1. **Use the model of demand and supply to explain the effects of third-party payers on the health-care market and on health-care spending.**

There has been much discussion over the past three decades about the health-care problem in the United States. Much of this discussion has focused on rising spending for health care. In this section, we will apply the model of demand and supply to health care to see what we can learn about some of the reasons behind rising spending in this important sector of the economy.

Figure 4.14 shows the share of U.S. output devoted to health care since 1960. In 1960, about 5% of total output was devoted to health care; by 2004 this share had risen to 15.4%. That has meant that we are devoting more of our spending to health care, and less to other goods and services, than we would be had health-care spending not risen so much.

FIGURE 4.14 Health-Care Spending as a Percentage of U.S. Output, 1960–2003

Health care's share of total U.S. output rose from about 5% in 1960 to 15.3% in 2003.



Data for period 1960–1992 from Health Care Finance Association (which was the predecessor to the Centers for Medicare and Medicaid Services); Data for period 1993–2003 from Centers for Medicare and Medicaid Services, Office of the Actuary: National Health Statistics Group http://www.cms.hhs.gov/statistics/nhe/historical/t1.asp.

Why were Americans willing to increase their spending on health care so dramatically? The model of demand and supply gives us part of the answer. As we apply the model to this problem, we will also gain a better understanding of the role of prices in a market economy.

3.1 The Demand and Supply for Health Care

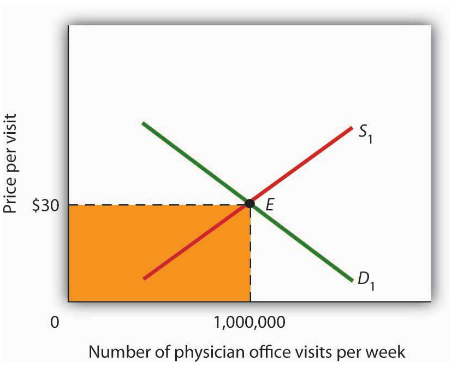
When we speak of "health care," we are speaking of the entire health-care industry. This industry produces services ranging from heart transplant operations to therapeutic massages; it produces goods ranging from X-ray machines to aspirin tablets. Clearly each of these goods and services is exchanged in a particular market. To assess the market forces affecting health care, we will focus first on just one of these markets: the market for physician office visits. When you go to the doctor, you are part of the demand for these visits. Your doctor, by seeing you, is part of the supply.

Figure 4.15 shows the market, assuming that it operates in a fashion similar to other markets. The demand curve D_1 and the supply curve S_1 intersect at point E, with an equilibrium price of $30 per office visit. The equilibrium quantity of office visits per week is 1,000,000.

We can use the demand and supply graph to show total spending, which equals the price per unit (in this case, $30 per visit) times the quantity consumed (in this case, 1,000,000 visits per week). Total spending for physician office visits thus equals $30,000,000 per week ($30 times 1,000,000 visits). We show total spending as the area of a rectangle bounded by the price and the quantity. It is the shaded region in Figure 4.15.

FIGURE 4.15 Total Spending for Physician Office Visits

Total spending on physician office visits is $30 per visit multiplied by 1,000,000 visits per week, which equals $30,000,000. It is the shaded area bounded by price and quantity.



The picture in Figure 4.15 misses a crucial feature of the market. Most people in the United States have health insurance, provided either by private firms, by private purchases, or by the government. With health insurance, people agree to pay a fixed amount to the insurer in exchange for the insurer's agreement to pay for most of the health-care expenses they incur. While insurance plans differ in their specific provisions, let us suppose that all individuals have plans that require them to pay $10 for an office visit; the insurance company will pay the rest.

How will this insurance affect the market for physician office visits? If it costs only $10 for a visit instead of $30, people will visit their doctors more often. The quantity of office visits demanded will increase. In Figure 4.16, this is shown as a movement along the demand curve. Think about your own choices. When you get a cold, do you go to the doctor? Probably not, if it is a minor cold. But if you feel like you are dying, or wish you were, you probably head for the doctor. Clearly, there are lots of colds in between these two extremes. Whether you drag yourself to the doctor will depend on the severity of your cold and what you will pay for a visit. At a lower price, you are more likely to go to the doctor; at a higher price, you are less likely to go.

In the case shown, the quantity of office visits rises to 1,500,000 per week. But that suggests a potential problem. The quantity of visits supplied at a price of $30 per visit was 1,000,000. According to supply curve S_1, it will take a price of $50 per visit to increase the quantity supplied to 1,500,000 visits (Point F on S_1). But consumers—patients—pay only $10.

third-party payer

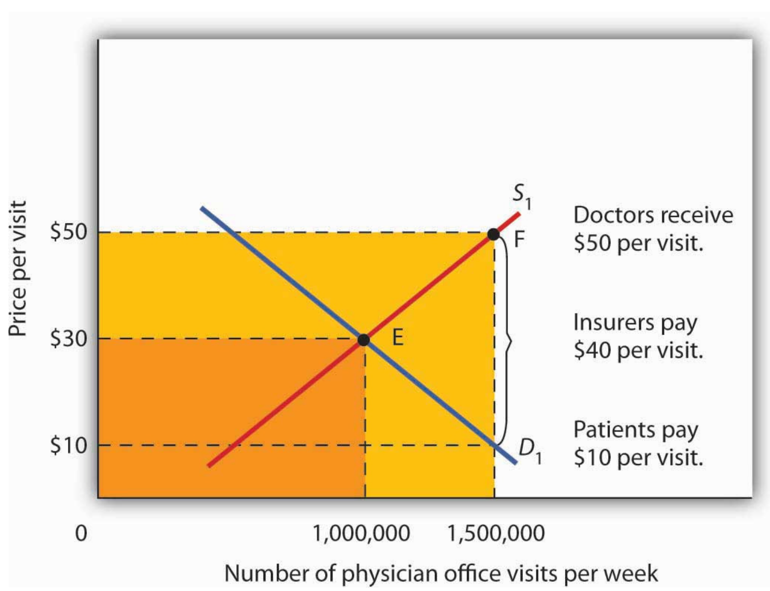
An agent other than the seller or the buyer who pays part of the price of a good or service.

Insurers make up the difference between the fees doctors receive and the price patients pay. In our example, insurers pay $40 per visit of insured patients to supplement the $10 that patients pay. When an agent other than the seller or the buyer pays part of the price of a good or service, we say that the agent is a **third-party payer**.

Notice how the presence of a third-party payer affects total spending on office visits. When people paid for their own visits, and the price equaled $30 per visit, total spending equaled $30 million per week. Now doctors receive $50 per visit and provide 1,500,000 visits per week. Total spending has risen to $75 million per week ($50 times 1,500,000 visits, shown by the darkly shaded region plus the lightly shaded region).

FIGURE 4.16 Total Spending for Physician Office Visits Covered by Insurance

With insurance, the quantity of physician office visits demanded rises to 1,500,000. The supply curve shows that it takes a price of $50 per visit to increase the quantity supplied to 1,500,000 visits. Patients pay $10 per visit and insurance pays $40 per visit. Total spending rises to $75,000,000 per week, shown by the darkly shaded region plus the lightly shaded region.



The response described in Figure 4.16 holds for many different types of goods and services covered by insurance or otherwise paid for by third-party payers. For example, the availability of scholarships and subsidized tuition at public and private universities increases the quantity of education demanded and the total expenditures on higher education. In markets with third-party payers, an equilibrium is achieved, but it is not at the intersection of the demand and supply curves. The effect of third-party payers is to decrease the price that consumers directly pay for the goods and services they consume and to increase the price that suppliers receive. Consumers use more than they would in the absence of third-party payers, and providers are encouraged to supply more than they otherwise would. The result is increased total spending.

KEY TAKEAWAYS

- The rising share of the output of the United States devoted to health care represents a rising opportunity cost. More spending on health care means less spending on other goods and services, compared to what would have transpired had health-care spending not risen so much.
- The model of demand and supply can be used to show the effect of third-party payers on total spending. With third-party payers (for example, health insurers), the quantity of services consumed rises, as does spending.

TRY IT!

The provision of university education through taxpayer-supported state universities is another example of a market with a third-party payer. Use the model of demand and supply to discuss the impact this has on the higher education market. Specifically, draw a graph similar to Figure 4.16. How would you label the axes? Show the equilibrium price and quantity in the absence of a third-party payer and indicate total spending on education. Now show the impact of lower tuition As a result of state support for education. How much education do students demand at the lower tuition? How much tuition must educational institutions receive to produce that much education? How much spending on education will occur? Compare total spending before and after a third-party payer enters this market.

Case in Point: The Oregon Plan

The health-care industry presents us with a dilemma. Clearly, it makes sense for people to have health insurance. Just as clearly, health insurance generates a substantial increase in spending for health care. If that spending is to be limited, some mechanism must be chosen to do it. One mechanism would be to require patients to pay a larger share of their own health-care consumption directly, reducing the payments made by third-party payers. Allowing people to accumulate tax-free private medical savings accounts is one way to do this. Another option is to continue the current trend to use insurance companies as the agents that limit spending. A third option is government regulation; this Case in Point describes how the state of Oregon tried to limit health-care spending by essentially refusing to be a third-party payer for certain services.

Like all other states, Oregon has wrestled with the problem of soaring Medicaid costs. Its solution to the problem illustrates some of the choices society might make in seeking to reduce health-care costs.

Oregon used to have a plan similar to plans in many other states. Households whose incomes were lower than 50% of the poverty line qualified for Medicaid. In 1987, the state began an effort to manage its Medicaid costs. It decided that it would no longer fund organ transplants and that it would use the money saved to give better care to pregnant women. The decision turned out to be a painful one; the first year, a seven-year-old boy with leukemia, who might have been saved with a bone marrow transplant, died. But state officials argued that the shift of expenditures to pregnant women would ultimately save more lives.

The state gradually expanded its concept of determining what services to fund and what services not to fund. It collapsed a list of 10,000 different diagnoses that had been submitted to its Medicaid program in the past into a list of more than 700 condition-treatment pairs. One such pair, for example, is appendicitis-appendectomy. Health-care officials then ranked these pairs in order of priority. The rankings were based on such factors as the seriousness of a particular condition and the cost and efficacy of treatments. The state announced that it would provide Medicaid to all households below the poverty line, but that it would not fund any procedure ranked below a certain level, initially number 588 on its list. The plan also set a budget limit for any one year; if spending rose above that limit, the legislature must appropriate additional money or drop additional procedures from the list of those covered by the plan. The Oregon Health Plan officially began operation in 1994.

While the Oregon plan has been applied only to households below the poverty line that are not covered by other programs, it suggests a means of reducing health-care spending. Clearly, if part of the health-care problem is excessive provision of services, a system designed to cut services must determine what treatments not to fund.

Professors Jonathan Oberlander, Theodore Marmor, and Lawrence Jacobs studied the impact of this plan in practice through the year 2000 and found that, in contrast to initial expectations, excluded procedures were generally ones of marginal medical value, so the "line in the sand" had little practical significance. In addition, they found that patients were often able to receive supposedly excluded services when physicians, for example, treated an uncovered illness in conjunction with a covered one. During the period of the study, the number of people covered by the plan expanded substantially and yet rationing of services essentially did not occur. How do they explain this seeming contradiction? Quite simply: state government increased revenues from various sources to support the plan. Indeed, they argue that, because treatments that might not be included were explicitly stated, political pressure made excluding them even more difficult and may have inadvertently increased the cost of the program.

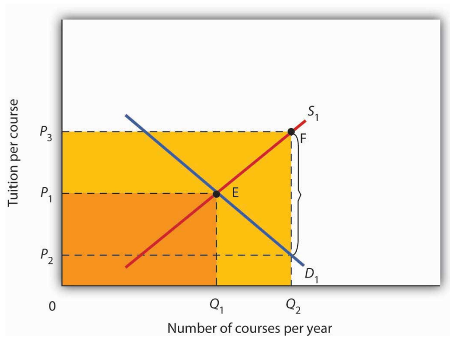
In the early 2000s, Oregon, like many other states, confronted severe budgetary pressures. To limit spending, it chose the perhaps less visible strategy of reducing the number of people covered through the plan. Once serving more than 100,000 people, budget cuts reduced the number served to about 17,000. Whereas in 1996, 11% of Oregonians lacked health insurance, in 2008 16% did.

Trailblazing again, in 2008 Oregon realized that its budget allowed room for coverage for a few thousand additional people. But how to choose among the 130,000 eligibles? The solution: to hold a lottery. More than 90,000 people queued up, hoping to be lucky winners.

Sources: Jonathan Oberlander, Theodore Marmor, and Lawrence Jacobs, "Rationing Medical Care: Rhetoric and Reality in the Oregon Health Plan," Canadian Medical Association Journal 164: 11 (May 29, 2001): 1583–1587; William Yardley, "Drawing Lots for Health Care," The New York Times, March 13, 2008: p. A12.

ANSWER TO TRY IT! PROBLEM

Without a third-party payer for education, the graph shows equilibrium tuition of P_1 and equilibrium quantity of education of Q_1. State support for education lowers tuition that students pay to P_2. As a result, students demand Q_2 courses per year. To provide that amount of education, educational institutions require tuition per course of P_3. Without a third-party payer, spending on education is $0P_1EQ_1$. With a third-party payer, spending rises to $0P_3FQ_2$.



4. REVIEW AND PRACTICE

Summary

In this chapter we used the tools of demand and supply to understand a wide variety of market outcomes. We learned that technological change and the entry of new sellers has caused the supply curve of personal computers to shift markedly to the right, thereby reducing equilibrium price and increasing equilibrium quantity. Market forces have made personal computers a common item in offices and homes.

Crude oil and gasoline prices soared in 2008 and then fell back. We looked at the causes of these increases as well as their impacts. Crude oil prices rose in large part As a result of increased demand, particularly from China. Higher prices for crude oil led to higher prices for gasoline. Those higher prices not only hurt consumers of gasoline, they also put upward pressure on the prices of a wide range of goods and services. Crude oil and gasoline prices then decreased dramatically in the last part of 2008, as world growth declined.

The model of demand and supply also explains the determination of stock prices. The price per share of corporate stock reflects the market's estimate of the expected profitability of the firm. Any information about the firm that causes potential buyers or current owners of corporate stock to reevaluate how profitable they think the firm is, or will be, will cause the equilibrium price of the stock to change.

We then examined markets in which some form of government price control keeps price permanently above or below equilibrium. A price floor leads to persistent surpluses because it is set above the equilibrium price, whereas a price ceiling, because it is set below the equilibrium price, leads to persistent shortages. We saw that interfering with the market mechanism may solve one problem but often creates other problems at the same time. We discussed what some of these unintended consequences might be. For example, agricultural price floors aimed at boosting farm income have also raised prices for consumers and cost taxpayers dearly, and the bulk of government payments have gone to large farms. Rent controls have lowered rents, but they have also reduced the quantity of rental housing supplied, created shortages, and sometimes led to various forms of "backdoor" payments, which sometimes force the price of rental housing above what would exist in the absence of controls.

Finally, we looked at the market for health care and a special feature behind demand and supply in this market that helps to explain why the share of output of the United States that is devoted to health care has risen. Health care is an example of a market in which there are third-party payers (primarily private insurers and the government). With third-party payers the quantity of health-care services consumed rises, as does health-care spending.

CONCEPT PROBLEMS

1. Like personal computers, digital cameras have become a common household item. Digital camera prices have plunged in the last 10 years. Use the model of demand and supply to explain the fall in price and increase in quantity.
2. Enron Corp. was one of several corporations convicted of fraud in its accounting practices during the early part of this decade. It had created dummy corporations to hide massive borrowing and to give it the appearance of extraordinary profitability. Use the model of demand and supply to explain the likely impact of such convictions on the stocks of other corporations.
3. During World War II there was a freeze on wages, and corporations found they could evade the freeze by providing other fringe benefits such as retirement funds for their employees. The Office of Price Administration, which administered the wage freeze, ruled that the offer of retirement funds was not a violation of the freeze. The Internal Revenue Service went along with this and ruled that employer-financed retirement plans were not taxable income. Was the wage freeze an example of a price floor or a price ceiling? Use the model of demand and supply to explain why employers began to offer such benefits to their employees.
4. The text argues that political instability in potential suppliers of oil such as Iraq and Venezuela accounts for a relatively steep supply curve for crude oil such as the one shown in Figure 4.2 Suppose that this instability eases considerably and that the world supply curve for crude oil becomes much flatter. Draw such a curve, and explain its implications for the world economy and for typical consumers.
5. Suppose that technological change affects the dairy industry in the same way it has affected the computer industry. However, suppose that dairy price supports remain in place. How would this affect government spending on the dairy program? Use the model of demand and supply to support your answer.
6. People often argue that there is a “shortage” of child care. Using the model of demand and supply, evaluate whether this argument is likely to be correct.
7. “During most of the past 50 years the United States has had a surplus of farmers, and this has been the root of the farm problem.” Comment.
8. Suppose the Department of Agriculture ordered all farmers to reduce the acreage they plant by 10%. Would you expect a 10% reduction in food production? Why or why not?
9. The text argues that the increase in gasoline prices had a particularly strong impact on low-income people. Name some other goods and services for which a sharp increase in price would have a similar impact on people with low incomes.
10. Suppose that the United States and the European Union impose a price ceiling on crude oil of $25 per barrel. Explain, and illustrate graphically, how this would affect the markets for crude oil and for gasoline in the United States and in the European Union.
11. Given that rent controls can actually hurt low-income people, devise a housing strategy that would provide affordable housing for those whose incomes fall below the poverty line (in 2004, this was about $19,000 for a family of four).
12. Using the model of demand and supply, show and explain how an increase in the share individuals must pay directly for medical care affects the quantity they consume. Explain how this would address the total amount of spending on health care.
13. Given that people pay premiums for their health insurance, how can we say that insurance lowers the prices people pay for health-care services?
14. Suppose that physicians now charge $30 for an office visit and insurance policies require patients to pay 33 1/3% of the amount they pay the physicians, so the out-of-pocket cost to consumers is $10 per visit. In an effort to control costs, the government imposes a price ceiling of $27 per office visit. Using a demand and supply model, show how this policy would affect the market for health care.
15. Do you think the U.S. health-care system requires reform? Why or why not? If you think reform is in order, explain the approach to reform you advocate.

NUMERICAL PROBLEMS

Problems 1–4 are based on the following demand and supply schedules for corn (all quantities are in millions of bushels per year).

Price per bushel	Quantity demanded	Quantity supplied
$0	6	0
1	5	1
2	4	2
3	3	3
4	2	4
5	1	5
6	0	6

1. Draw the demand and supply curves for corn. What is the equilibrium price? The equilibrium quantity?
2. Suppose the government now imposes a price floor at $4 per bushel. Show the effect of this program graphically. How large is the surplus of corn?
3. With the price floor, how much do farmers receive for their corn? How much would they have received if there were no price floor?
4. If the government buys all the surplus wheat, how much will it spend?

Problems 5–9 are based on the following hypothetical demand and supply curves for apartments

Rent/Month	Number of Apts. Demanded/Month	Number of Apts. Supplied/Month
$0	120,000	0
200	100,000	20,000
400	80,000	40,000
600	60,000	60,000
800	40,000	80,000
1000	20,000	100,000
1200	0	120,000

5. Draw the demand and supply curves for apartments.
6. What is the equilibrium rent per month? At this rent, what is the number of apartments demanded and supplied per month?
7. Suppose a ceiling on rents is set at $400 per month. Characterize the situation that results from this policy.
8. At the rent ceiling, how many apartments are demanded? How many are supplied?
9. How much are people willing to pay for the number of apartments supplied at the ceiling? Describe the arrangements to which this situation might lead.

ENDNOTES

1. Ilkka Tuomi, "The Lives and Death of Moore's Law." http://www.firstmonday.org/issues/issue7_11/tuomi/index. *First Monday* (http://www.firstmonday.org) is a peer-reviewed journal on the Internet.
2. Richard Arnott, "Time for Revisionism on Rent Control," *Journal of Economic Perspectives* 9(1) (Winter, 1995): 99–120.

CHAPTER 5

Elasticity: A Measure of Response

START UP: RAISE FARES? LOWER FARES? WHAT'S A PUBLIC TRANSIT MANAGER TO DO?

Imagine that you are the manager of the public transportation system for a large metropolitan area. Operating costs for the system have soared in the last few years, and you are under pressure to boost revenues. What do you do?

An obvious choice would be to raise fares. That will make your customers angry, but at least it will generate the extra revenue you need—or will it? The law of demand says that raising fares will reduce the number of passengers riding on your system. If the number of passengers falls only a little, then the higher fares that your remaining passengers are paying might produce the higher revenues you need. But what if the number of passengers falls by so much that your higher fares actually reduce your revenues? If that happens, you will have made your customers mad and your financial problem worse!

Maybe you should recommend *lower* fares. After all, the law of demand also says that lower fares will increase the number of passengers. Having more people use the public transportation system could more than offset a lower fare you collect from each person. But it might not. What *will* you do?

Your job and the fiscal health of the public transit system are riding on your making the correct decision. To do so, you need to know just how responsive the quantity demanded is to a price change. You need a measure of responsiveness.

Economists use a measure of responsiveness called elasticity. **Elasticity** is the ratio of the percentage change in a dependent variable to a percentage change in an independent variable. If the dependent variable is *y*, and the independent variable is *x*, then the elasticity of *y* with respect to a change in *x* is given by:

elasticity

The ratio of the percentage change in a dependent variable to a percentage change in an independent variable.

$$e_{y,\,x} = \frac{\%\text{ change in }y}{\%\text{ change in }x}$$

A variable such as *y* is said to be more elastic (responsive) if the percentage change in *y* is large relative to the percentage change in *x*. It is less elastic if the reverse is true.

As manager of the public transit system, for example, you will want to know how responsive the number of passengers on your system (the dependent variable) will be to a change in fares (the independent variable). The concept of elasticity will help you solve your public transit pricing problem and a great many other issues in economics. We will examine several elasticities in this chapter—all will tell us how responsive one variable is to a change in another.

1. THE PRICE ELASTICITY OF DEMAND

LEARNING OBJECTIVES

1. **Explain the concept of price elasticity of demand and its calculation.**
2. **Explain what it means for demand to be price inelastic, unit price elastic, price elastic, perfectly price inelastic, and perfectly price elastic.**
3. **Explain how and why the value of the price elasticity of demand changes along a linear demand curve.**
4. **Understand the relationship between total revenue and price elasticity of demand.**
5. **Discuss the determinants of price elasticity of demand.**

We know from the law of demand how the quantity demanded will respond to a price change: it will change in the opposite direction. But how *much* will it change? It seems reasonable to expect, for example, that a 10% change in the price charged for a visit to the doctor would yield a different percentage change in quantity demanded than a 10% change in the price of a Ford Mustang. But how much is this difference?

price elasticity of demand

The percentage change in quantity demanded of a particular good or service divided by the percentage change in the price of that good or service, all other things unchanged.

To show how responsive quantity demanded is to a change in price, we apply the concept of elasticity. The **price elasticity of demand** for a good or service, e_D, is the percentage change in quantity demanded of a particular good or service divided by the percentage change in the price of that good or service, all other things unchanged. Thus we can write

EQUATION 5.2

$$e_D = \frac{\text{\% change in quantity demanded}}{\text{\% change in price}}$$

Because the price elasticity of demand shows the responsiveness of quantity demanded to a price change, assuming that other factors that influence demand are unchanged, it reflects movements *along* a demand curve. With a downward-sloping demand curve, price and quantity demanded move in opposite directions, so the price elasticity of demand is always negative. A positive percentage change in price implies a negative percentage change in quantity demanded, and vice versa. Sometimes you will see the absolute value of the price elasticity measure reported. In essence, the minus sign is ignored because it is expected that there will be a negative (inverse) relationship between quantity demanded and price. In this text, however, we will retain the minus sign in reporting price elasticity of demand and will say "the absolute value of the price elasticity of demand" when that is what we are describing.

Heads Up!

Be careful not to confuse elasticity with slope. The slope of a line is the change in the value of the variable on the vertical axis divided by the change in the value of the variable on the horizontal axis between two points. Elasticity is the ratio of the percentage changes. The slope of a demand curve, for example, is the ratio of the change in price to the change in quantity between two points on the curve. The price elasticity of demand is the ratio of the percentage change in quantity to the percentage change in price. As we will see, when computing elasticity at different points on a linear demand curve, the slope is constant—that is, it does not change—but the value for elasticity will change.

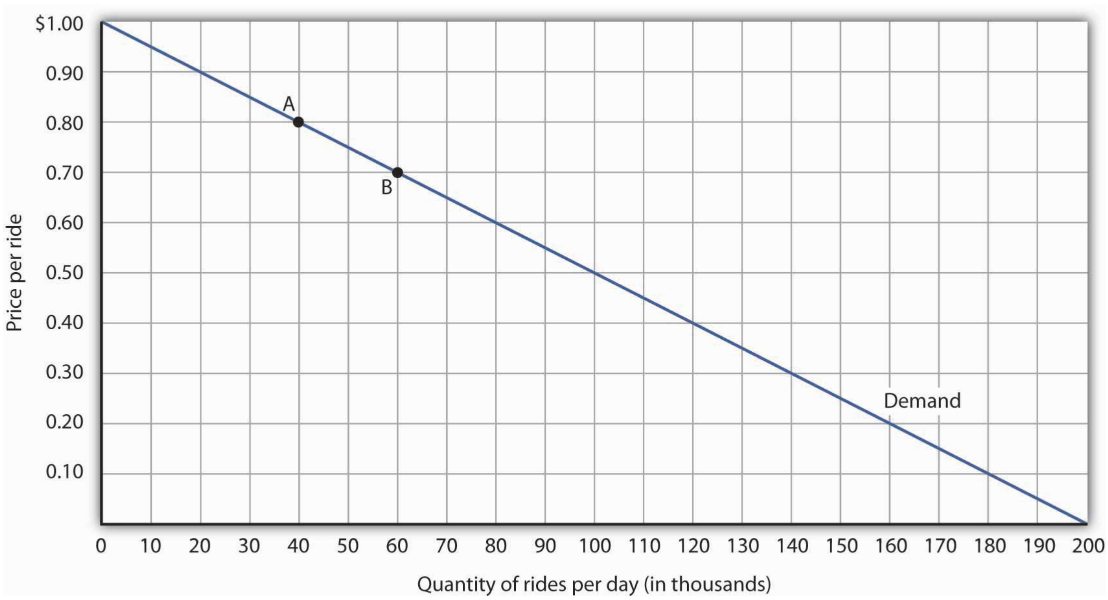
1.1 Computing the Price Elasticity of Demand

Finding the price elasticity of demand requires that we first compute percentage changes in price and in quantity demanded. We calculate those changes between two points on a demand curve.

Figure 5.1 shows a particular demand curve, a linear demand curve for public transit rides. Suppose the initial price is \$0.80, and the quantity demanded is 40,000 rides per day; we are at point A on the curve. Now suppose the price falls to \$0.70, and we want to report the responsiveness of the quantity demanded. We see that at the new price, the quantity demanded rises to 60,000 rides per day (point B). To compute the elasticity, we need to compute the percentage changes in price and in quantity demanded between points A and B.

FIGURE 5.1 **Responsiveness and Demand**

The demand curve shows how changes in price lead to changes in the quantity demanded. A movement from point A to point B shows that a \$0.10 reduction in price increases the number of rides per day by 20,000. A movement from B to A is a \$0.10 increase in price, which reduces quantity demanded by 20,000 rides per day.



We measure the percentage change between two points as the change in the variable divided by the *average* value of the variable between the two points. Thus, the percentage change in quantity between points A and B in Figure 5.1 is computed relative to the *average* of the quantity values at points A and B: (60,000 + 40,000)/2 = 50,000. The percentage change in quantity, then, is 20,000/50,000, or 40%. Likewise, the percentage change in price between points A and B is based on the *average* of the two prices: (\$0.80+\$0.70)/2=\$0.75, and so we have a percentage change of −0.10/0.75, or −13.33%. The price elasticity of demand between points A and B is thus 40%/(−13.33%)=−3.00.

This measure of elasticity, which is based on percentage changes relative to the average value of each variable between two points, is called **arc elasticity**. The arc elasticity method has the advantage that it yields the same elasticity whether we go from point A to point B or from point B to point A. It is the method we shall use to compute elasticity.

arc elasticity

Measure of elasticity based on percentage changes relative to the average value of each variable between two points.

For the arc elasticity method, we calculate the price elasticity of demand using the average value of price, $\bar{P}$, and the average value of quantity demanded, $\bar{Q}$. We shall use the Greek letter Δ to mean "change in," so the change in quantity between two points is ΔQ and the change in price is ΔP. Now we can write the formula for the price elasticity of demand as

EQUATION 5.3

$$e_D = \frac{\Delta Q \,/\, \bar{Q}}{\Delta P \,/\, \bar{P}}$$

The price elasticity of demand between points A and B is thus:

$$e_D = \frac{\frac{20{,}000}{(40{,}000 + 60{,}000)/2}}{\frac{-\$0.10}{(\$0.80 + \$0.70)/2}} = \frac{40\%}{-13.33\%} = -3.00$$

With the arc elasticity formula, the elasticity is the same whether we move from point A to point B or from point B to point A. If we start at point B and move to point A, we have:

$$e_D = \frac{\frac{-20{,}000}{(60{,}000 + 40{,}000)/2}}{\frac{0.10}{(\$0.70 + \$0.80)/2}} = \frac{-40\%}{13.33\%} = -3.00$$

The arc elasticity method gives us an estimate of elasticity. It gives the value of elasticity at the midpoint over a range of change, such as the movement between points A and B. For a precise computation of elasticity, we would need to consider the response of a dependent variable to an extremely small change in an independent variable. The fact that arc elasticities are approximate suggests an important practical rule in calculating arc elasticities: we should consider only small changes in independent variables. We cannot apply the concept of arc elasticity to large changes.

Another argument for considering only small changes in computing price elasticities of demand will become evident in the next section. We will investigate what happens to price elasticities as we move from one point to another along a linear demand curve.

Heads Up!

Notice that in the arc elasticity formula, the method for computing a percentage change differs from the standard method with which you may be familiar. That method measures the percentage change in a variable relative to its original value. For example, using the standard method, when we go from point A to point B, we would compute the percentage change in quantity as 20,000/40,000=50%. The percentage change in price would be −\$0.10/\$0.80=−12.5%. The price elasticity of demand would then be 50%/(−12.5%)=−4.00. Going from point B to point A, however, would yield a different elasticity. The percentage change in quantity would be −20,000/60,000, or −33.33%. The percentage change in price would be \$0.10/\$0.70=14.29%. The price elasticity of demand would thus be −33.33%/14.29%=−2.33. By using the average quantity and average price to calculate percentage changes, the arc elasticity approach avoids the necessity to specify the direction of the change and, thereby, gives us the same answer whether we go from A to B or from B to A.

1.2 Price Elasticities Along a Linear Demand Curve

What happens to the price elasticity of demand when we travel along the demand curve? The answer depends on the nature of the demand curve itself. On a linear demand curve, such as the one in Figure 5.2, elasticity becomes smaller (in absolute value) as we travel downward and to the right.

FIGURE 5.2 Price Elasticities of Demand for a Linear Demand Curve

The price elasticity of demand varies between different pairs of points along a linear demand curve. The lower the price and the greater the quantity demanded, the lower the absolute value of the price elasticity of demand.

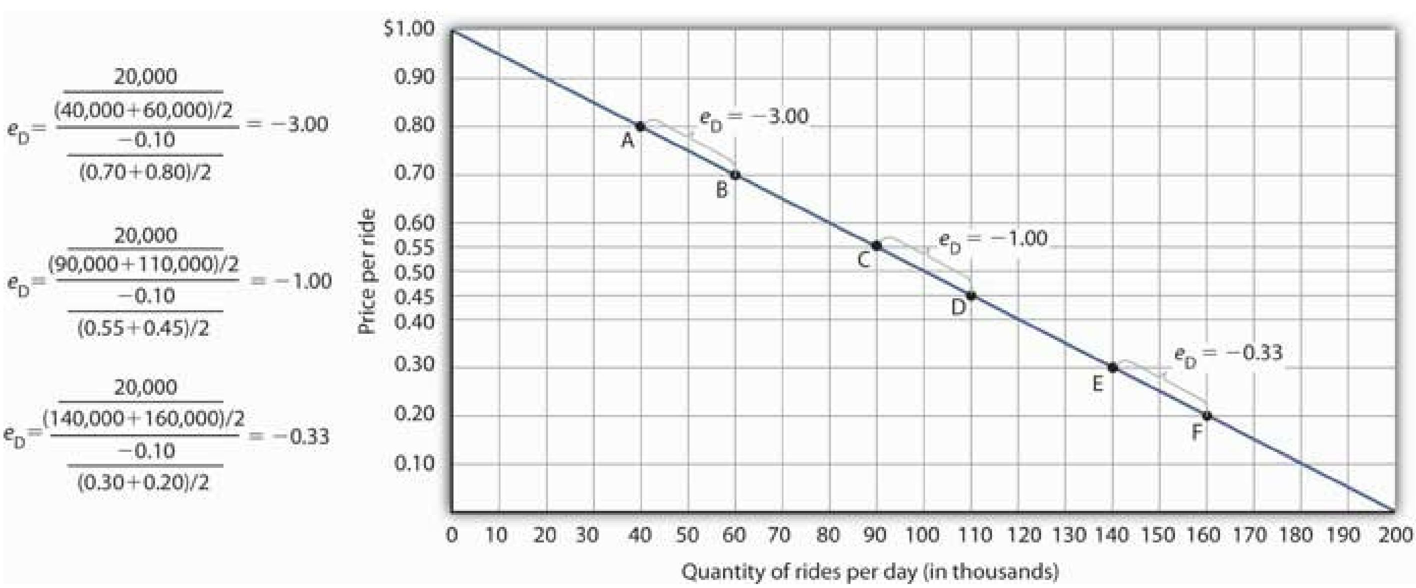


Figure 5.2 shows the same demand curve we saw in Figure 5.1. We have already calculated the price elasticity of demand between points A and B; it equals −3.00. Notice, however, that when we use the same method to compute the price elasticity of demand between other sets of points, our answer varies. For each of the pairs of points shown, the changes in price and quantity demanded are the same (a \$0.10 decrease in price and 20,000 additional rides per day, respectively). But at the high prices and low quantities on the upper part of the demand curve, the percentage change in quantity is relatively large, whereas the percentage change in price is relatively small. The absolute value of the price elasticity of demand is thus relatively large. As we move down the demand curve, equal changes in quantity represent smaller and smaller percentage changes, whereas equal changes in price represent larger and larger percentage changes, and the absolute value of the elasticity measure declines. Between points C and D,

for example, the price elasticity of demand is −1.00, and between points E and F the price elasticity of demand is −0.33.

On a linear demand curve, the price elasticity of demand varies depending on the interval over which we are measuring it. For any linear demand curve, the absolute value of the price elasticity of demand will fall as we move down and to the right along the curve.

1.3 The Price Elasticity of Demand and Changes in Total Revenue

Suppose the public transit authority is considering raising fares. Will its total revenues go up or down? **Total revenue** is the price per unit times the number of units sold.[1] In this case, it is the fare times the number of riders. The transit authority will certainly want to know whether a price increase will cause its total revenue to rise or fall. In fact, determining the impact of a price change on total revenue is crucial to the analysis of many problems in economics.

total revenue

A firm's output multiplied by the price at which it sells that output.

We will do two quick calculations before generalizing the principle involved. Given the demand curve shown in Figure 5.2, we see that at a price of $0.80, the transit authority will sell 40,000 rides per day. Total revenue would be $32,000 per day ($0.80 times 40,000). If the price were lowered by $0.10 to $0.70, quantity demanded would increase to 60,000 rides and total revenue would increase to $42,000 ($0.70 times 60,000). The reduction in fare *increases* total revenue. However, if the initial price had been $0.30 and the transit authority reduced it by $0.10 to $0.20, total revenue would *decrease* from $42,000 ($0.30 times 140,000) to $32,000 ($0.20 times 160,000). So it appears that the impact of a price change on total revenue depends on the initial price and, by implication, the original elasticity. We generalize this point in the remainder of this section.

The problem in assessing the impact of a price change on total revenue of a good or service is that a change in price always changes the quantity demanded in the opposite direction. An increase in price reduces the quantity demanded, and a reduction in price increases the quantity demanded. The question is how much. Because total revenue is found by multiplying the price per unit times the quantity demanded, it is not clear whether a change in price will cause total revenue to rise or fall.

We have already made this point in the context of the transit authority. Consider the following three examples of price increases for gasoline, pizza, and diet cola.

Suppose that 1,000 gallons of gasoline per day are demanded at a price of $4.00 per gallon. Total revenue for gasoline thus equals $4,000 per day (=1,000 gallons per day times $4.00 per gallon). If an increase in the price of gasoline to $4.25 reduces the quantity demanded to 950 gallons per day, total revenue rises to $4,037.50 per day (=950 gallons per day times $4.25 per gallon). Even though people consume less gasoline at $4.25 than at $4.00, total revenue rises because the higher price more than makes up for the drop in consumption.

Next consider pizza. Suppose 1,000 pizzas per week are demanded at a price of $9 per pizza. Total revenue for pizza equals $9,000 per week (=1,000 pizzas per week times $9 per pizza). If an increase in the price of pizza to $10 per pizza reduces quantity demanded to 900 pizzas per week, total revenue will still be $9,000 per week (=900 pizzas per week times $10 per pizza). Again, when price goes up, consumers buy less, but this time there is no change in total revenue.

Now consider diet cola. Suppose 1,000 cans of diet cola per day are demanded at a price of $0.50 per can. Total revenue for diet cola equals $500 per day (=1,000 cans per day times $0.50 per can). If an increase in the price of diet cola to $0.55 per can reduces quantity demanded to 880 cans per month, total revenue for diet cola falls to $484 per day (=880 cans per day times $0.55 per can). As in the case of gasoline, people will buy less diet cola when the price rises from $0.50 to $0.55, but in this example total revenue drops.

In our first example, an increase in price increased total revenue. In the second, a price increase left total revenue unchanged. In the third example, the price rise reduced total revenue. Is there a way to predict how a price change will affect total revenue? There is; the effect depends on the price elasticity of demand.

Elastic, Unit Elastic, and Inelastic Demand

price elastic

Situation in which the absolute value of the price elasticity of demand is greater than 1.

unit price elastic

Situation in which the absolute value of the price elasticity of demand is equal to 1.

price inelastic

Situation in which the absolute value of the price of elasticity of demand is less than 1.

To determine how a price change will affect total revenue, economists place price elasticities of demand in three categories, based on their absolute value. If the absolute value of the price elasticity of demand is greater than 1, demand is termed **price elastic**. If it is equal to 1, demand is **unit price elastic**. And if it is less than 1, demand is **price inelastic**.

Relating Elasticity to Changes in Total Revenue

When the price of a good or service changes, the quantity demanded changes in the opposite direction. Total revenue will move in the direction of the variable that changes by the larger percentage. If the variables move by the same percentage, total revenue stays the same. If quantity demanded changes by a larger percentage than price (i.e., if demand is price elastic), total revenue will change in the direction of the quantity change. If price changes by a larger percentage than quantity demanded (i.e., if demand is price inelastic), total revenue will move in the direction of the price change. If price and quantity demanded change by the same percentage (i.e., if demand is unit price elastic), then total revenue does not change.

When demand is price inelastic, a given percentage change in price results in a smaller percentage change in quantity demanded. That implies that total revenue will move in the direction of the price change: a reduction in price will reduce total revenue, and an increase in price will increase it.

Consider the price elasticity of demand for gasoline. In the example above, 1,000 gallons of gasoline were purchased each day at a price of $4.00 per gallon; an increase in price to $4.25 per gallon reduced the quantity demanded to 950 gallons per day. We thus had an average quantity of 975 gallons per day and an average price of $4.125. We can thus calculate the arc price elasticity of demand for gasoline:

Percentage change in quantity demanded = −50 / 975 = −5.1%
Percentage change in price = 0.25 / 4.125=6.06%
Price elasticity of demand = −5.1% / 6.06% = −0.84

The demand for gasoline is price inelastic, and total revenue moves in the direction of the price change. When price rises, total revenue rises. Recall that in our example above, total spending on gasoline (which equals total revenues to sellers) rose from $4,000 per day (=1,000 gallons per day times $4.00) to $4037.50 per day (=950 gallons per day times $4.25 per gallon).

When demand is price inelastic, a given percentage change in price results in a smaller percentage change in quantity demanded. That implies that total revenue will move in the direction of the price change: an increase in price will increase total revenue, and a reduction in price will reduce it.

Consider again the example of pizza that we examined above. At a price of $9 per pizza, 1,000 pizzas per week were demanded. Total revenue was $9,000 per week (=1,000 pizzas per week times $9 per pizza). When the price rose to $10, the quantity demanded fell to 900 pizzas per week. Total revenue remained $9,000 per week (=900 pizzas per week times $10 per pizza). Again, we have an average quantity of 950 pizzas per week and an average price of $9.50. Using the arc elasticity method, we can compute:

Percentage change in quantity demanded = − 100 / 950 = − 10.5%
Percentage change in price = $1.00 / $9.50 = 10.5%
Price elasticity of demand = − 10.5% / 10.5% = − 1.0

Demand is unit price elastic, and total revenue remains unchanged. Quantity demanded falls by the same percentage by which price increases.

Consider next the example of diet cola demand. At a price of $0.50 per can, 1,000 cans of diet cola were purchased each day. Total revenue was thus $500 per day (=$0.50 per can times 1,000 cans per day). An increase in price to $0.55 reduced the quantity demanded to 880 cans per day. We thus have an average quantity of 940 cans per day and an average price of $0.525 per can. Computing the price elasticity of demand for diet cola in this example, we have:

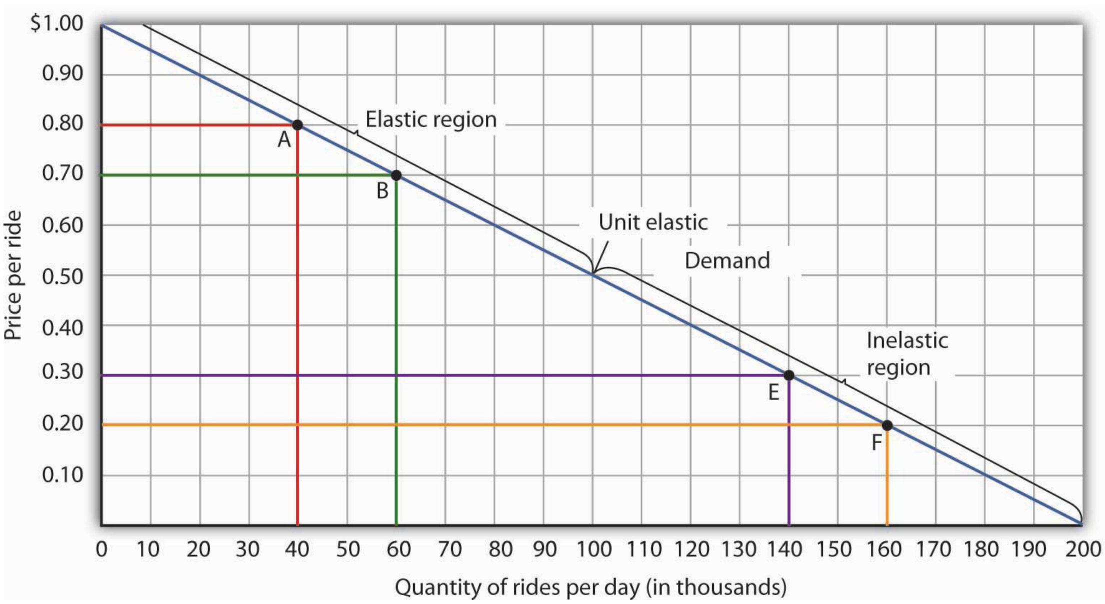
Percentage change in quantity demanded = − 120 / 940 = − 12.8%
Percentage change in price = $0.05 / $0.525 = 9.5%
Price elasticity of demand = − 12.8% / 9.5% = − 1.3

The demand for diet cola is price elastic, so total revenue moves in the direction of the quantity change. It falls from $500 per day before the price increase to $484 per day after the price increase.

A demand curve can also be used to show changes in total revenue. Figure 5.3 shows the demand curve from Figure 5.1 and Figure 5.2. At point A, total revenue from public transit rides is given by the area of a rectangle drawn with point A in the upper right-hand corner and the origin in the lower left-hand corner. The height of the rectangle is price; its width is quantity. We have already seen that total revenue at point A is \$32,000 (\$0.80 × 40,000). When we reduce the price and move to point B, the rectangle showing total revenue becomes shorter and wider. Notice that the area gained in moving to the rectangle at B is greater than the area lost; total revenue rises to \$42,000 (\$0.70 × 60,000). Recall from Figure 5.2 that demand is elastic between points A and B. In general, demand is elastic in the upper half of any linear demand curve, so total revenue moves in the direction of the quantity change.

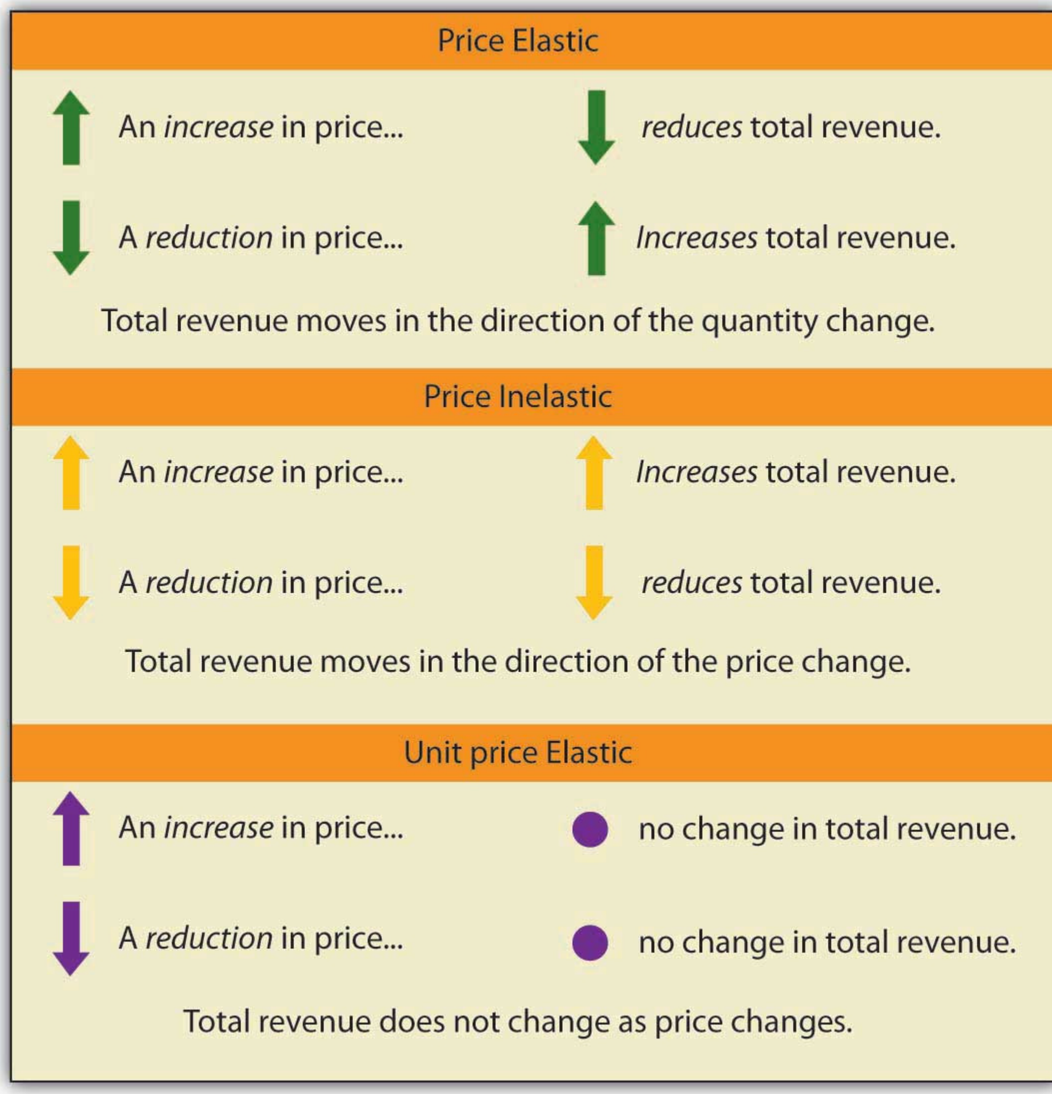
FIGURE 5.3 Changes in Total Revenue and a Linear Demand Curve

Moving from point A to point B implies a reduction in price and an increase in the quantity demanded. Demand is elastic between these two points. Total revenue, shown by the areas of the rectangles drawn from points A and B to the origin, rises. When we move from point E to point F, which is in the inelastic region of the demand curve, total revenue falls.



A movement from point E to point F also shows a reduction in price and an increase in quantity demanded. This time, however, we are in an inelastic region of the demand curve. Total revenue now moves in the direction of the price change—it falls. Notice that the rectangle drawn from point F is smaller in area than the rectangle drawn from point E, once again confirming our earlier calculation.

FIGURE 5.4



We have noted that a linear demand curve is more elastic where prices are relatively high and quantities relatively low and less elastic where prices are relatively low and quantities relatively high. We can be even more specific. *For any linear demand curve, demand will be price elastic in the upper half of the curve and price inelastic in its lower half. At the midpoint of a linear demand curve, demand is unit price elastic.*

1.4 Constant Price Elasticity of Demand Curves

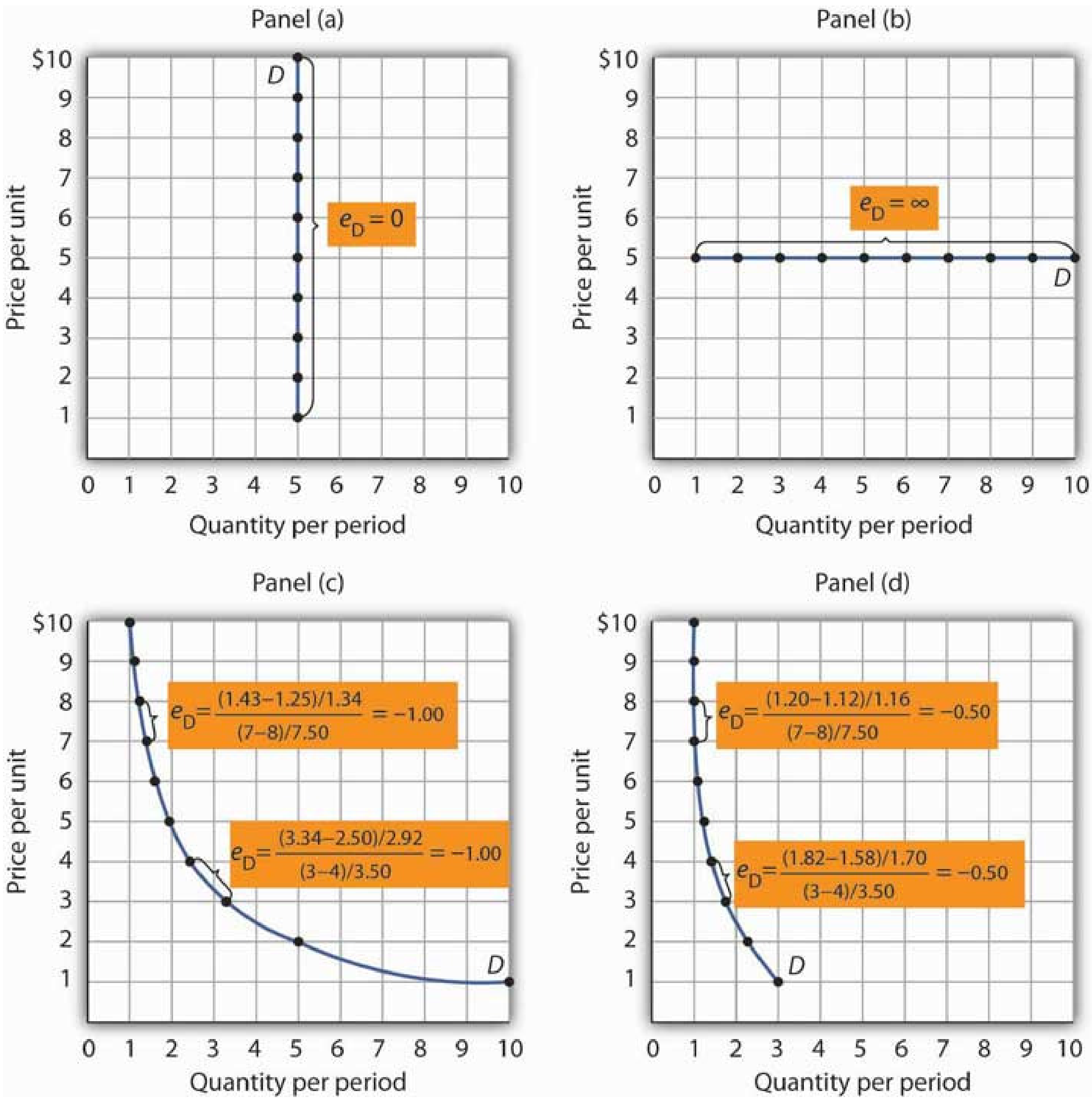
perfectly inelastic

Situation in which the price elasticity of demand is zero.

Figure 5.5 shows four demand curves over which price elasticity of demand is the same at all points. The demand curve in Panel (a) is vertical. This means that price changes have no effect on quantity demanded. The numerator of the formula given in Equation 5.2 for the price elasticity of demand (percentage change in quantity demanded) is zero. The price elasticity of demand in this case is therefore zero, and the demand curve is said to be **perfectly inelastic**. This is a theoretically extreme case, and no good that has been studied empirically exactly fits it. A good that comes close, at least over a specific price range, is insulin. A diabetic will not consume more insulin as its price falls but, over some price range, will consume the amount needed to control the disease.

FIGURE 5.5 **Demand Curves with Constant Price Elasticities**

The demand curve in Panel (a) is perfectly inelastic. The demand curve in Panel (b) is perfectly elastic. Price elasticity of demand is –1.00 all along the demand curve in Panel (c), whereas it is –0.50 all along the demand curve in Panel (d).



As illustrated in Figure 5.5, several other types of demand curves have the same elasticity at every point on them. The demand curve in Panel (b) is horizontal. This means that even the smallest price changes have enormous effects on quantity demanded. The denominator of the formula given in Equation 5.2 for the price elasticity of demand (percentage change in price) approaches zero. The price elasticity of demand in this case is therefore infinite, and the demand curve is said to be **perfectly elastic**.[2] This is the type of demand curve faced by producers of standardized products such as wheat. If the wheat of other farms is selling at $4 per bushel, a typical farm can sell as much wheat as it wants to at $4 but nothing at a higher price and would have no reason to offer its wheat at a lower price.

perfectly elastic

Situation in which the price elasticity of demand is infinite.

The nonlinear demand curves in Panels (c) and (d) have price elasticities of demand that are negative; but, unlike the linear demand curve discussed above, the value of the price elasticity is constant all along each demand curve. The demand curve in Panel (c) has price elasticity of demand equal to –1.00 throughout its range; in Panel (d) the price elasticity of demand is equal to –0.50 throughout its range. Empirical estimates of demand often show curves like those in Panels (c) and (d) that have the same elasticity at every point on the curve.

Heads Up!

Do not confuse price inelastic demand and perfectly inelastic demand. Perfectly inelastic demand means that the change in quantity is zero for any percentage change in price; the demand curve in this case is vertical. Price inelastic demand means only that the percentage change in quantity is less than the percentage change in price, not that the change in quantity is zero. With price inelastic (as opposed to perfectly inelastic) demand, the demand curve itself is still downward sloping.

1.5 Determinants of the Price Elasticity of Demand

The greater the absolute value of the price elasticity of demand, the greater the responsiveness of quantity demanded to a price change. What determines whether demand is more or less price elastic? The most important determinants of the price elasticity of demand for a good or service are the availability of substitutes, the importance of the item in household budgets, and time.

Availability of Substitutes

The price elasticity of demand for a good or service will be greater in absolute value if many close substitutes are available for it. If there are lots of substitutes for a particular good or service, then it is easy for consumers to switch to those substitutes when there is a price increase for that good or service. Suppose, for example, that the price of Ford automobiles goes up. There are many close substitutes for Fords—Chevrolets, Chryslers, Toyotas, and so on. The availability of close substitutes tends to make the demand for Fords more price elastic.

If a good has no close substitutes, its demand is likely to be somewhat less price elastic. There are no close substitutes for gasoline, for example. The price elasticity of demand for gasoline in the intermediate term of, say, three–nine months is generally estimated to be about −0.5. Since the absolute value of price elasticity is less than 1, it is price inelastic. We would expect, though, that the demand for a particular brand of gasoline will be much more price elastic than the demand for gasoline in general.

Importance in Household Budgets

One reason price changes affect quantity demanded is that they change how much a consumer can buy; a change in the price of a good or service affects the purchasing power of a consumer's income and thus affects the amount of a good the consumer will buy. This effect is stronger when a good or service is important in a typical household's budget.

A change in the price of jeans, for example, is probably more important in your budget than a change in the price of pencils. Suppose the prices of both were to double. You had planned to buy four pairs of jeans this year, but now you might decide to make do with two new pairs. A change in pencil prices, in contrast, might lead to very little reduction in quantity demanded simply because pencils are not likely to loom large in household budgets. The greater the importance of an item in household budgets, the greater the absolute value of the price elasticity of demand is likely to be.

Time

Suppose the price of electricity rises tomorrow morning. What will happen to the quantity demanded?

The answer depends in large part on how much time we allow for a response. If we are interested in the reduction in quantity demanded by tomorrow afternoon, we can expect that the response will be very small. But if we give consumers a year to respond to the price change, we can expect the response to be much greater. We expect that the absolute value of the price elasticity of demand will be greater when more time is allowed for consumer responses.

Consider the price elasticity of crude oil demand. Economist John C. B. Cooper estimated short- and long-run price elasticities of demand for crude oil for 23 industrialized nations for the period 1971–2000. Professor Cooper found that for virtually every country, the price elasticities were negative, and the long-run price elasticities were generally much greater (in absolute value) than were the short-run price elasticities. His results are reported in Table 5.1. As you can see, the research was reported in a journal published by OPEC (Organization of Petroleum Exporting Countries), an organization whose members have profited greatly from the inelasticity of demand for their product. By restricting supply, OPEC, which produces about 45% of the world's crude oil, is able to put upward pressure on the price of crude. That increases OPEC's (and all other oil producers') total revenues and reduces total costs.

TABLE 5.1 Short- and Long-Run Price Elasticities of the Demand for Crude Oil in 23 Countries

For most countries, price elasticity of demand for crude oil tends to be greater (in absolute value) in the long run than in the short run.

Country	Short-Run Price Elasticity of Demand	Long-Run Price Elasticity of Demand
Australia	–0.034	–0.068
Austria	–0.059	–0.092
Canada	–0.041	–0.352
China	0.001	0.005
Denmark	–0.026	–0.191
Finland	–0.016	–0.033
France	–0.069	–0.568
Germany	–0.024	–0.279
Greece	–0.055	–0.126
Iceland	–0.109	–0.452
Ireland	–0.082	–0.196
Italy	–0.035	–0.208
Japan	–0.071	–0.357
Korea	–0.094	–0.178
Netherlands	–0.057	–0.244
New Zealand	–0.054	–0.326
Norway	–0.026	–0.036
Portugal	0.023	0.038
Spain	–0.087	–0.146
Sweden	–0.043	–0.289
Switzerland	–0.030	–0.056
United Kingdom	–0.068	–0.182
United States	–0.061	–0.453

Source: John C. B. Cooper, "Price Elasticity of Demand for Crude Oil: Estimates from 23 Countries," OPEC Review: Energy Economics & Related Issues, 27:1 (March 2003): 4. The estimates are based on data for the period 1971–2000, except for China and South Korea, where the period is 1979–2000. While the price elasticities for China and Portugal were positive, they were not statistically significant.

KEY TAKEAWAYS

- The price elasticity of demand measures the responsiveness of quantity demanded to changes in price; it is calculated by dividing the percentage change in quantity demanded by the percentage change in price.
- Demand is price inelastic if the absolute value of the price elasticity of demand is less than 1; it is unit price elastic if the absolute value is equal to 1; and it is price elastic if the absolute value is greater than 1.
- Demand is price elastic in the upper half of any linear demand curve and price inelastic in the lower half. It is unit price elastic at the midpoint.
- When demand is price inelastic, total revenue moves in the direction of a price change. When demand is unit price elastic, total revenue does not change in response to a price change. When demand is price elastic, total revenue moves in the direction of a quantity change.
- The absolute value of the price elasticity of demand is greater when substitutes are available, when the good is important in household budgets, and when buyers have more time to adjust to changes in the price of the good.

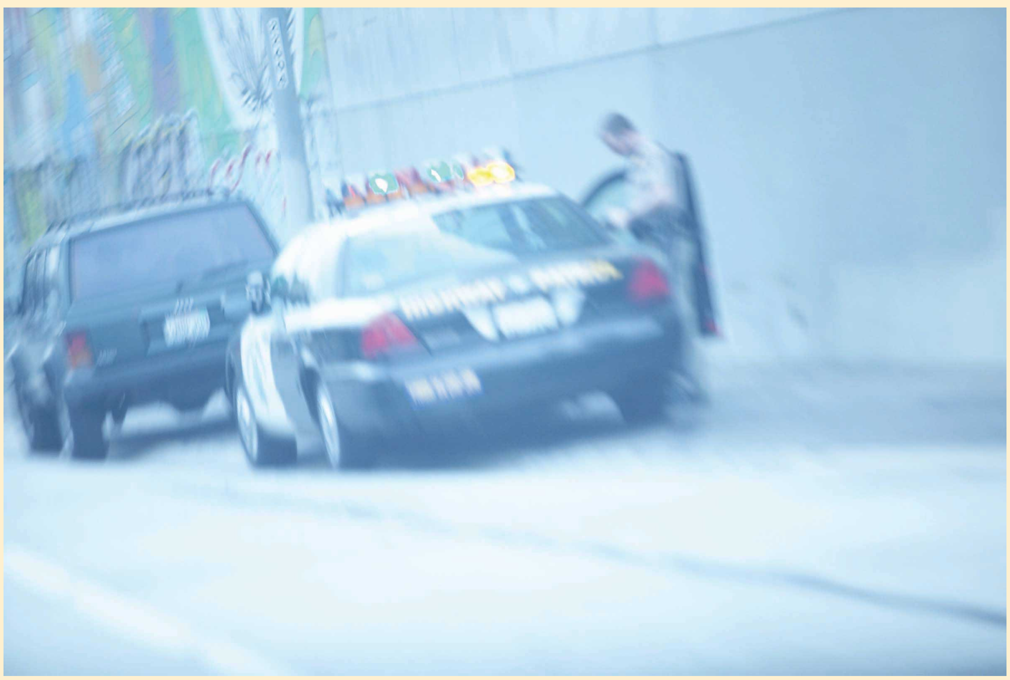
TRY IT!

You are now ready to play the part of the manager of the public transit system. Your finance officer has just advised you that the system faces a deficit. Your board does not want you to cut service, which means that you cannot cut costs. Your only hope is to increase revenue. Would a fare increase boost revenue?

You consult the economist on your staff who has researched studies on public transportation elasticities. She reports that the estimated price elasticity of demand for the first few months after a price change is about −0.3, but that after several years, it will be about −1.5.

1. Explain why the estimated values for price elasticity of demand differ.
2. Compute what will happen to ridership and revenue over the next few months if you decide to raise fares by 5%.
3. Compute what will happen to ridership and revenue over the next few years if you decide to raise fares by 5%.
4. What happens to total revenue now and after several years if you choose to raise fares?

Case in Point: Elasticity and Stop Lights



We all face the situation every day. You are approaching an intersection. The yellow light comes on. You know that you are supposed to slow down, but you are in a bit of a hurry. So, you speed up a little to try to make the light. But the red light flashes on just before you get to the intersection. Should you risk it and go through?

Many people faced with that situation take the risky choice. In 1998, 2,000 people in the United States died as a result of drivers running red lights at intersections. In an effort to reduce the number of drivers who make such choices, many areas have installed cameras at intersections. Drivers who run red lights have their pictures taken and receive citations in the mail. This enforcement method, together with recent increases in the fines for driving through red lights at intersections, has led to an intriguing application of the concept of elasticity. Economists Avner Bar-Ilan of the University of Haifa in Israel and Bruce Sacerdote of Dartmouth University have estimated what is, in effect, the price elasticity for driving through stoplights with respect to traffic fines at intersections in Israel and in San Francisco.

In December 1996, Israel sharply increased the fine for driving through a red light. The old fine of 400 shekels (this was equal at that time to $122 in the United States) was increased to 1,000 shekels ($305). In January 1998, California raised its fine for the offense from $104 to $271. The country of Israel and the city of San Francisco installed cameras at several intersections. Drivers who ignored stoplights got their pictures taken and automatically received citations imposing the new higher fines.

We can think of driving through red lights as an activity for which there is a demand—after all, ignoring a red light speeds up one's trip. It may also generate satisfaction to people who enjoy disobeying traffic laws. The concept of elasticity gives us a way to show just how responsive drivers were to the increase in fines.

Professors Bar-Ilan and Sacerdote obtained information on all the drivers cited at 73 intersections in Israel and eight intersections in San Francisco. For Israel, for example, they defined the period January 1992 to June 1996 as the "before" period. They compared the number of violations during the before period to the number of violations from July 1996 to December 1999—the "after" period—and found there was a reduction in tickets per driver of 31.5 per cent. Specifically, the average number of tickets per driver was 0.073 during the period before the increase; it fell to 0.050 after the increase. The increase in the fine was 150 per cent. (Note that, because they were making a "before" and "after" calculation, the authors used the standard method described in the Heads Up! on computing a percentage change—i.e., they computed the percentage changes in comparison to the original values instead of the average value of the variables.) The elasticity of citations with respect to the fine was thus −0.21 (= −31.5%/150%).

The economists estimated elasticities for particular groups of people. For example, young people (age 17–30) had an elasticity of −0.36; people over the age of 30 had an elasticity of −0.16. In general, elasticities fell in absolute value as income rose. For San Francisco and Israel combined, the elasticity was between −0.26 and −0.33.

In general, the results showed that people responded rationally to the increases in fines. Increasing the price of a particular behavior reduced the frequency of that behavior. The study also points out the effectiveness of cameras as an enforcement technique. With cameras, violators can be certain they will be cited if they ignore a red light. And reducing the number of people running red lights clearly saves lives.

Source: Avner Bar-Ilan and Bruce Sacerdote. "The Response of Criminals and Non-Criminals to Fines." Journal of Law and Economics, 47:1 (April 2004): 1–17.

ANSWERS TO TRY IT! PROBLEMS

1. The absolute value of price elasticity of demand tends to be greater when more time is allowed for consumers to respond. Over time, riders of the commuter rail system can organize car pools, move, or otherwise adjust to the fare increase.
2. Using the formula for price elasticity of demand and plugging in values for the estimate of price elasticity (−0.5) and the percentage change in price (5%) and then rearranging terms, we can solve for the percentage change in quantity demanded as: $e_D = \%\Delta \text{ in } Q/\%\Delta \text{ in } P$; $-0.5 = \%\Delta \text{ in } Q/5\%$; $(-0.5)(5\%) = \%\Delta \text{ in } Q = -2.5\%$. Ridership falls by 2.5% in the first few months.
3. Using the formula for price elasticity of demand and plugging in values for the estimate of price elasticity over a few years (−1.5) and the percentage change in price (5%), we can solve for the percentage change in quantity demanded as $e_D = \%\Delta \text{ in } Q/\%\Delta \text{ in } P$; $-1.5 = \%\Delta \text{ in } Q/5\%$; $(-1.5)(5\%) = \%\Delta \text{ in } Q = -7.5\%$. Ridership falls by 7.5% over a few years.
4. Total revenue rises immediately after the fare increase, since demand over the immediate period is price inelastic. Total revenue falls after a few years, since demand changes and becomes price elastic.

2. RESPONSIVENESS OF DEMAND TO OTHER FACTORS

LEARNING OBJECTIVES

1. **Explain the concept of income elasticity of demand and its calculation.**
2. **Classify goods as normal or inferior depending on their income elasticity of demand.**
3. **Explain the concept of cross price elasticity of demand and its calculation.**
4. **Classify goods as substitutes or complements depending on their cross price elasticity of demand.**

Although the response of quantity demanded to changes in price is the most widely used measure of elasticity, economists are interested in the response to changes in the demand shifters as well. Two of the most important measures show how demand responds to changes in income and to changes in the prices of related goods and services.

2.1 Income Elasticity of Demand

income elasticity of demand

The percentage change in quantity demanded at a specific price divided by the percentage change in income that produced the demand change, all other things unchanged.

We saw in the chapter that introduced the model of demand and supply that the demand for a good or service is affected by income. We measure the **income elasticity of demand**, e_Y, as the percentage change in quantity demanded *at a specific price* divided by the percentage change in income that produced the demand change, all other things unchanged:

EQUATION 5.4

$$e_Y = \frac{\%\text{ change in quantity demanded}}{\%\text{ change in income}}$$

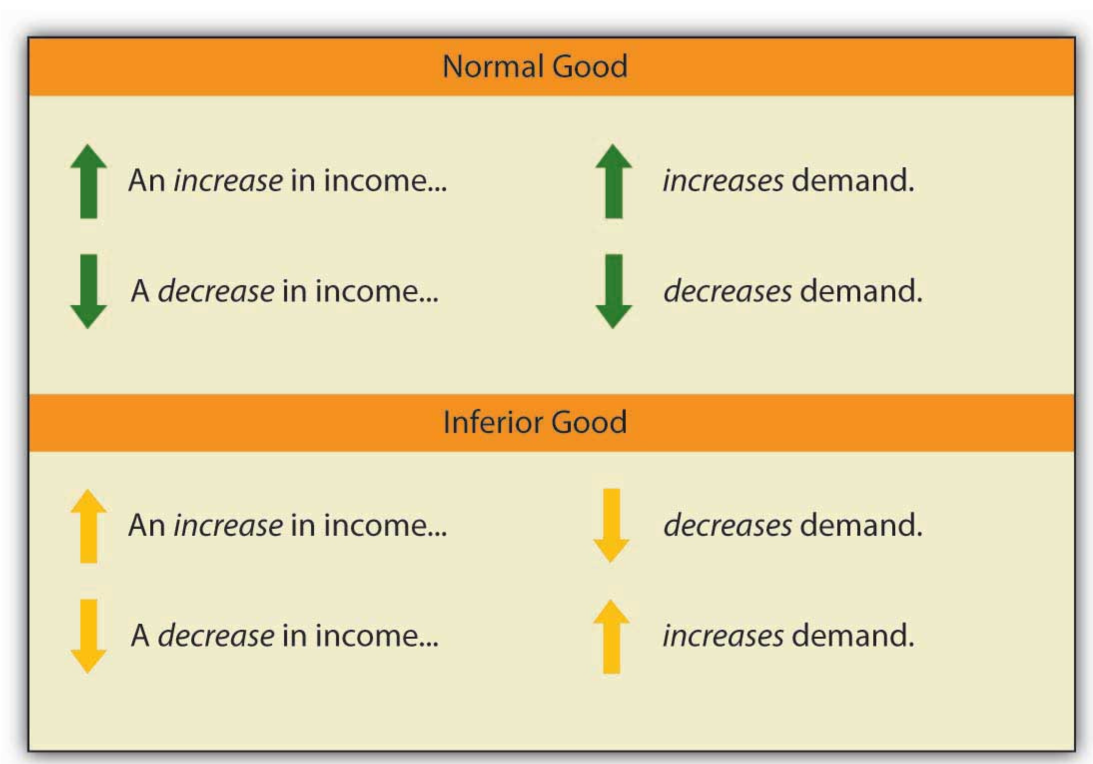
The symbol *Y* is often used in economics to represent income. Because income elasticity of demand reports the responsiveness of quantity demanded to a change in income, all other things unchanged (including the price of the good), it reflects a shift in the demand curve at a given price. Remember that price elasticity of demand reflects movements *along* a demand curve in response to a change in price.

A positive income elasticity of demand means that income and demand move in the same direction—an increase in income increases demand, and a reduction in income reduces demand. As we learned, a good whose demand rises as income rises is called a normal good.

Studies show that most goods and services are normal, and thus their income elasticities are positive. Goods and services for which demand is likely to move in the same direction as income include housing, seafood, rock concerts, and medical services.

If a good or service is inferior, then an increase in income reduces demand for the good. That implies a negative income elasticity of demand. Goods and services for which the income elasticity of demand is likely to be negative include used clothing, beans, and urban public transit. For example, the studies we have already cited concerning the demands for urban public transit in France and in Madrid found the long-run income elasticities of demand to be negative (−0.23 in France and −0.25 in Madrid).[3]

FIGURE 5.7



When we compute the income elasticity of demand, we are looking at the change in the quantity demanded at a specific price. We are thus dealing with a change that shifts the demand curve. An increase in income shifts the demand for a normal good to the right; it shifts the demand for an inferior good to the left.

2.2 Cross Price Elasticity of Demand

The demand for a good or service is affected by the prices of related goods or services. A reduction in the price of salsa, for example, would increase the demand for chips, suggesting that salsa is a complement of chips. A reduction in the price of chips, however, would reduce the demand for peanuts, suggesting that chips are a substitute for peanuts.

The measure economists use to describe the responsiveness of demand for a good or service to a change in the price of another good or service is called the **cross price elasticity of demand**, $e_{A, B}$. It equals the percentage change in the quantity demanded of one good or service *at a specific price* divided by the percentage change in the price of a related good or service. We are varying the price of a related good when we consider the cross price elasticity of demand, so the response of quantity demanded is shown as a shift in the demand curve.

cross price elasticity of demand

It equals the percentage change in the quantity demanded of one good or service at a specific price divided by the percentage change in the price of a related good or service.

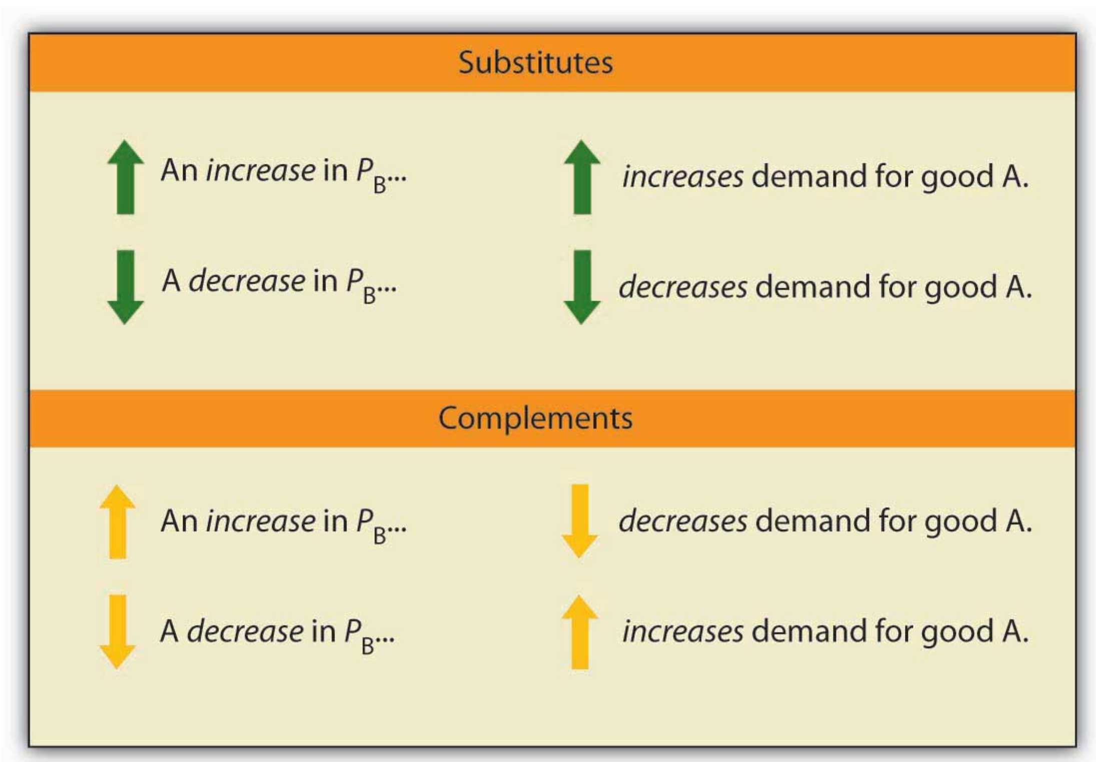
The cross price elasticity of the demand for good A with respect to the price of good B is given by:

EQUATION 5.5

$$e_{A, B} = \frac{\text{\% change in quantity demanded of good A}}{\text{\% change in price of good B}}$$

Cross price elasticities of demand define whether two goods are substitutes, complements, or unrelated. If two goods are substitutes, an increase in the price of one will lead to an increase in the demand for the other—the cross price elasticity of demand is positive. If two goods are complements, an increase in the price of one will lead to a reduction in the demand for the other—the cross price elasticity of demand is negative. If two goods are unrelated, a change in the price of one will not affect the demand for the other—the cross price elasticity of demand is zero.

FIGURE 5.8



An examination of the demand for local television advertising with respect to the price of local radio advertising revealed that the two goods are clearly substitutes. A 10 per cent increase in the price of local radio advertising led to a 10 per cent increase in demand for local television advertising, so that the cross price elasticity of demand for local television advertising with respect to changes in the price of radio advertising was 1.0.[4]

Heads Up!

Notice that with income elasticity of demand and cross price elasticity of demand we are primarily concerned with whether the measured value of these elasticities is positive or negative. In the case of income elasticity of demand this tells us whether the good or service is normal or inferior. In the case of cross price elasticity of demand it tells us whether two goods are substitutes or complements. With price elasticity of demand we were concerned with whether the measured absolute value of this elasticity was greater than, less than, or equal to 1, because this gave us information about what happens to total revenue as price changes. The terms elastic and inelastic apply to price elasticity of demand. They are not used to describe income elasticity of demand or cross price elasticity of demand.

KEY TAKEAWAYS

- The income elasticity of demand reflects the responsiveness of demand to changes in income. It is the percentage change in quantity demanded *at a specific price* divided by the percentage change in income, ceteris paribus.
- Income elasticity is positive for normal goods and negative for inferior goods.
- The cross price elasticity of demand measures the way demand for one good or service responds to changes in the price of another. It is the percentage change in the quantity demanded of one good or service *at a specific price* divided by the percentage change in the price of another good or service, all other things unchanged.
- Cross price elasticity is positive for substitutes, negative for complements, and zero for goods or services whose demands are unrelated.

TRY IT!

Suppose that when the price of bagels rises by 10%, the demand for cream cheese falls by 3% at the current price, and that when income rises by 10%, the demand for bagels increases by 1% at the current price. Calculate the cross price elasticity of demand for cream cheese with respect to the price of bagels and tell whether bagels and cream cheese are substitutes or complements. Calculate the income elasticity of demand and tell whether bagels are normal or inferior.

Case in Point: Teen Smoking and Elasticity

© 2010 Jupiterimages Corporation

Tobacco kills more people than any other substance or disease. Worldwide, the annual death toll from tobacco is well over 3 million people per year. In the United States alone, 400,000 people die each year as a result of their use of tobacco.

More than two-thirds of smokers indicated in 1995 that they would prefer to quit smoking but were unable to do so, according to the Centers for Disease Control and Prevention. In fact, less than 2.5 per cent of smokers succeed in quitting each year.

Most smokers begin using tobacco as teenagers. Teens tend to underestimate the danger of smoking and to overestimate their likely ability to quit smoking when they choose to do so. One can, therefore, argue that the decision to smoke may not be a rational one, and it is one that imposes substantial costs on the rest of society. Because it raises health-care costs, it raises health insurance rates. And the evidence is mounting that secondhand smoke imposes serious health consequences on nonsmokers. Because smoking is such a serious problem for our society, and because the decision to smoke is typically made when one is a teenager, health economists tend to focus on measures to prevent young people from smoking.

One place to begin in limiting teen smoking is price. The price elasticity of demand for teenage smokers is greater (in absolute value) than that for the population in general because the cost of tobacco represents a greater percentage of teen incomes than of adult incomes. For all smokers, the price elasticity of demand was estimated by economists Matthew C. Farrelly, Terry F. Pechacek, and Frank J. Chaloupka to be −0.32. Health-care economists estimate that the price elasticity of demand for cigarettes for teenagers is between −0.9 and −1.5.

In 1998, the tobacco industry reached a settlement with 46 states that had filed lawsuits against the industry, charging that the tobacco industry had imposed huge health-care costs. The Master Settlement Agreement (MSA) called for a payment of $205 billion over a period of 25 years (the other four states reached separate agreements with the industry in 1997 and 1998). The MSA led to an increase in the price of cigarettes by 48 per cent between 1997 and 1999. The percentage of high school students who smoked fell significantly by 2000, indicating a substantial responsiveness of teenagers to price changes.

The MSA also required that states use some of the money they receive from tobacco firms to carry out antismoking programs. The nature and scope of these programs vary widely. State excise taxes, also varying widely, range from 2.5¢ per pack in Virginia (a tobacco-producing state) to $1.51 in Massachusetts. Given the greater responsiveness of teenagers to the price of cigarettes, excise taxes should prove an effective device.

One caveat, however, in evaluating the impact of a tax hike on teen smoking is that some teens might switch from cigarettes to smokeless tobacco, which is associated with a higher risk of oral cancer. It is estimated that for young males the cross price elasticity of smokeless tobacco with respect to the price of cigarettes is 1.2—a 10% increase in cigarette prices leads to a 12% increase in young males using smokeless tobacco.

Sources: Matthew C. Farrelly, Terry F. Pechacek, and Frank J. Chaloupka; "The Impact of Tobacco Control Program Expenditures on Aggregate Cigarette Sales: 1981–2000," Journal of Health Economics 22:5 (September 2003): 843–859; Michael Grossman, "Cigarette Taxes," Public Health Reports 112:4 (July/August 1997): 290–297; Hana Ross and Frank J. Chaloupka, "The Effect of Public Policies and Prices on Youth Smoking," Southern Economic Journal 70:4 (April 2004): 796–815; John A. Tauras. "Public Policy and Smoking Cessation among Young Adults in the United States," Health Policy, 68:3 (June 2004): 321–332.

ANSWER TO TRY IT! PROBLEM

Using the formula for cross price elasticity of demand, we find that $e_{AB} = (-3\%)/(10\%) = -0.3$. Since the e_{AB} is negative, bagels and cream cheese are complements. Using the formula for income elasticity of demand, we find that $e_Y = (+1\%)/(10\%) = +0.1$. Since e_Y is positive, bagels are a normal good.

3. PRICE ELASTICITY OF SUPPLY

LEARNING OBJECTIVES

1. **Explain the concept of elasticity of supply and its calculation.**
2. **Explain what it means for supply to be price inelastic, unit price elastic, price elastic, perfectly price inelastic, and perfectly price elastic.**
3. **Explain why time is an important determinant of price elasticity of supply.**
4. **Apply the concept of price elasticity of supply to the labor supply curve.**

The elasticity measures encountered so far in this chapter all relate to the demand side of the market. It is also useful to know how responsive quantity supplied is to a change in price.

Suppose the demand for apartments rises. There will be a shortage of apartments at the old level of apartment rents and pressure on rents to rise. All other things unchanged, the more responsive the quantity of apartments supplied is to changes in monthly rents, the lower the increase in rent required to eliminate the shortage and to bring the market back to equilibrium. Conversely, if quantity supplied is less responsive to price changes, price will have to rise more to eliminate a shortage caused by an increase in demand.

This is illustrated in Figure 5.10. Suppose the rent for a typical apartment had been R_0 and the quantity Q_0 when the demand curve was D_1 and the supply curve was either S_1 (a supply curve in which quantity supplied is less responsive to price changes) or S_2 (a supply curve in which quantity supplied is more responsive to price changes). Note that with either supply curve, equilibrium price and quantity are initially the same. Now suppose that demand increases to D_2, perhaps due to population growth. With supply curve S_1, the price (rent in this case) will rise to R_1 and the quantity of apartments will rise to Q_1. If, however, the supply curve had been S_2, the rent would only have to rise to R_2 to bring the market back to equilibrium. In addition, the new equilibrium number of apartments would be higher at Q_2. Supply curve S_2 shows greater responsiveness of quantity supplied to price change than does supply curve S_1.

We measure the **price elasticity of supply** (e_S) as the ratio of the percentage change in quantity supplied of a good or service to the percentage change in its price, all other things unchanged:

EQUATION 5.6

$$e_S = \frac{\text{\% change in quantity supplied}}{\text{\% change in price}}$$

Because price and quantity supplied usually move in the same direction, the price elasticity of supply is usually positive. The larger the price elasticity of supply, the more responsive the firms that supply the good or service are to a price change.

Supply is price elastic if the price elasticity of supply is greater than 1, unit price elastic if it is equal to 1, and price inelastic if it is less than 1. A vertical supply curve, as shown in Panel (a) of Figure 5.11, is perfectly inelastic; its price elasticity of supply is zero. The supply of Beatles' songs is perfectly inelastic because the band no longer exists. A horizontal supply curve, as shown in Panel (b) of Figure 5.11, is perfectly elastic; its price elasticity of supply is infinite. It means that suppliers are willing to supply any amount at a certain price.

FIGURE 5.10 Increase in Apartment Rents Depends on How Responsive Supply Is

The more responsive the supply of apartments is to changes in price (rent in this case), the less rents rise when the demand for apartments increases.

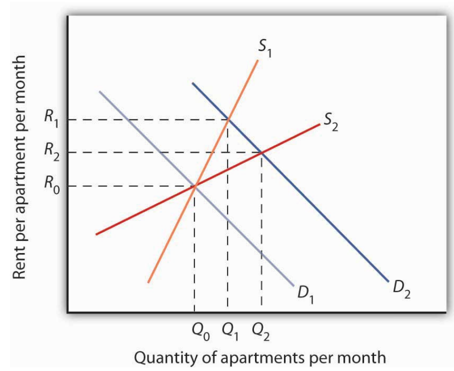
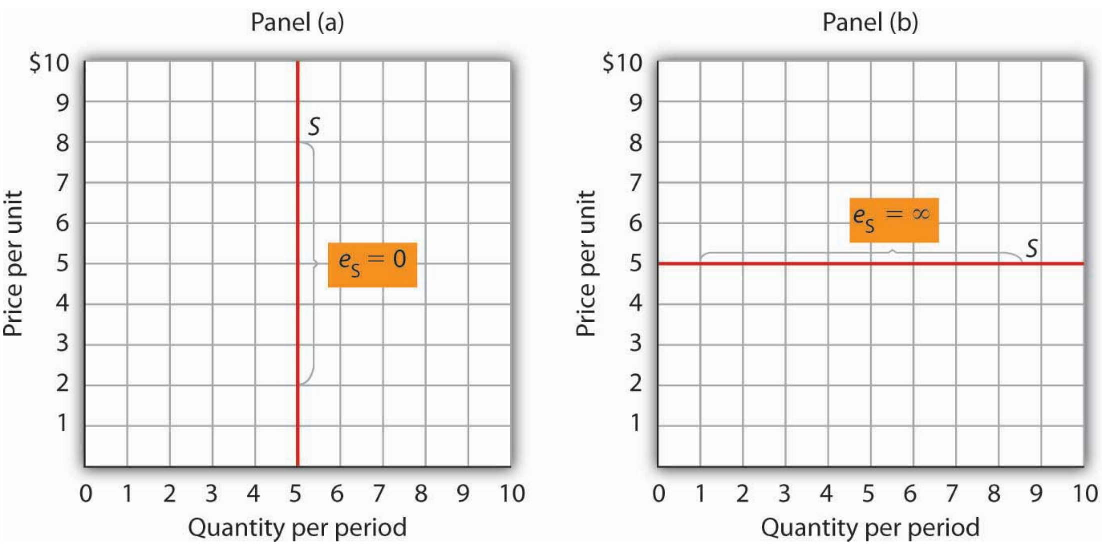


FIGURE 5.11 Supply Curves and Their Price Elasticities

The supply curve in Panel (a) is perfectly inelastic. In Panel (b), the supply curve is perfectly elastic.



price elasticity of supply

The ratio of the percentage change in quantity supplied of a good or service to the percentage change in its price, all other things unchanged.

3.1 Time: An Important Determinant of the Elasticity of Supply

Time plays a very important role in the determination of the price elasticity of supply. Look again at the effect of rent increases on the supply of apartments. Suppose apartment rents in a city rise. If we are looking at a supply curve of apartments over a period of a few months, the rent increase is likely to induce apartment owners to rent out a relatively small number of additional apartments. With the higher rents, apartment owners may be more vigorous in reducing their vacancy rates, and, indeed, with more people looking for apartments to rent, this should be fairly easy to accomplish. Attics and basements are easy to renovate and rent out as additional units. In a short period of time, however, the supply response is likely to be fairly modest, implying that the price elasticity of supply is fairly low. A supply curve corresponding to a short period of time would look like S_1 in Figure 5.10. It is during such periods that there may be calls for rent controls.

If the period of time under consideration is a few years rather than a few months, the supply curve is likely to be much more price elastic. Over time, buildings can be converted from other uses and new apartment complexes can be built. A supply curve corresponding to a longer period of time would look like S_2 in Figure 5.10.

3.2 Elasticity of Labor Supply: A Special Application

The concept of price elasticity of supply can be applied to labor to show how the quantity of labor supplied responds to changes in wages or salaries. What makes this case interesting is that it has sometimes been found that the measured elasticity is negative, that is, that an increase in the wage rate is associated with a reduction in the quantity of labor supplied.

In most cases, labor supply curves have their normal upward slope: higher wages induce people to work more. For them, having the additional income from working more is preferable to having more leisure time. However, wage increases may lead some people in very highly paid jobs to cut back on the number of hours they work because their incomes are already high and they would rather have more time for leisure activities. In this case, the labor supply curve would have a negative slope. The reasons for this phenomenon are explained more fully in a later chapter.

This chapter has covered a variety of elasticity measures. All report the degree to which a dependent variable responds to a change in an independent variable. As we have seen, the degree of this response can play a critically important role in determining the outcomes of a wide range of economic events. Table 5.2[5] provides examples of some estimates of elasticities.

TABLE 5.2 Selected Elasticity Estimates

Product	Elasticity	Product	Elasticity	Product	Elasticity
Price Elasticity of Demand		**Cross Price Elasticity of Demand**		**Income Elasticity of Demand**	
Crude oil (U.S.)*	−0.06	Alcohol with respect to price of heroin	−0.05	Speeding citations	−0.26 to −0.33
Gasoline	−0.1	Fuel with respect to price of transport	−0.48	Urban Public Trust in France and Madrid (respectively)	−0.23; −0.26
Speeding citations	−0.21	Alcohol with respect to price of food	−0.16	Ground beef	−0.197
Cabbage	−0.25	Marijuana with respect to price of heroin (similar for cocaine)	−0.01	Lottery instant game sales in Colorado	−0.06
Cocaine (two estimates)	−0.28; −1.0	Beer with respect to price of wine distilled liquor (young drinkers)	0.0	Heroin	−0.00
Alcohol	−0.30	Beer with respect to price of distilled liquor (young drinkers)	0.0	Marijuana, alcohol, cocaine	+0.00
Peaches	−0.38	Pork with respect to price of poultry	0.06	Potatoes	0.15
Marijuana	−0.4	Pork with respect to price of ground beef	0.23	Food**	0.2
Cigarettes (all smokers; two estimates)	−0.4; −0.32	Ground beef with respect to price of poultry	0.24	Clothing***	0.3
Crude oil (U.S.)**	−0.45	Ground beef with respect to price of pork	0.35	Beer	0.4
Milk (two estimates)	−0.49; −0.63	Coke with respect to price of Pepsi	0.61	Eggs	0.57
Gasoline (intermediate term)	−0.5	Pepsi with respect to price of Coke	0.80	Coke	0.60
Soft drinks	−0.55	Local television advertising with respect to price of radio advertising	1.0	Shelter**	0.7
Transportation*	−0.6	Smokeless tobacco with respect to price of cigarettes (young males)	1.2	Beef (table cuts—not ground)	0.81
Food	−0.7	**Price Elasticity of Supply**		Oranges	0.83
Beer	−0.7 to −0.9	Physicians (Specialist)	−0.3	Apples	1.32
Cigarettes (teenagers; two estimates)	−0.9 to −1.5	Physicians (Primary Care)	0.0	Leisure**	1.4
Heroin	−0.94	Physicians (Young male)	0.2	Peaches	1.43
Ground beef	−1.0	Physicians (Young female)	0.5	Health care**	1.6
Cottage cheese	−1.1	Milk*	0.36	Higher education	1.67
Gasoline**	−1.5	Milk**	0.5		
Coke	−1.71	Child care labor	2		
Transportation	−1.9				
Pepsi	−2.08				
Fresh tomatoes	−2.22				
Food**	−2.3				
Lettuce	−2.58				
Note: *=short-run; **=long-run					

KEY TAKEAWAYS

- The price elasticity of supply measures the responsiveness of quantity supplied to changes in price. It is the percentage change in quantity supplied divided by the percentage change in price. It is usually positive.
- Supply is price inelastic if the price elasticity of supply is less than 1; it is unit price elastic if the price elasticity of supply is equal to 1; and it is price elastic if the price elasticity of supply is greater than 1. A vertical supply curve is said to be perfectly inelastic. A horizontal supply curve is said to be perfectly elastic.
- The price elasticity of supply is greater when the length of time under consideration is longer because over time producers have more options for adjusting to the change in price.
- When applied to labor supply, the price elasticity of supply is usually positive but can be negative. If higher wages induce people to work more, the labor supply curve is upward sloping and the price elasticity of supply is positive. In some very high-paying professions, the labor supply curve may have a negative slope, which leads to a negative price elasticity of supply.

TRY IT!

In the late 1990s, it was reported on the news that the high-tech industry was worried about being able to find enough workers with computer-related expertise. Job offers for recent college graduates with degrees in computer science went with high salaries. It was also reported that more undergraduates than ever were majoring in computer science. Compare the price elasticity of supply of computer scientists at that point in time to the price elasticity of supply of computer scientists over a longer period of, say, 1999 to 2009.

Case in Point: A Variety of Labor Supply Elasticities

Studies support the idea that labor supply is less elastic in high-paying jobs than in lower-paying ones.

For example, David M. Blau estimated the labor supply of child-care workers to be very price elastic, with estimated price elasticity of labor supply of about 2.0. This means that a 10% increase in wages leads to a 20% increase in the quantity of labor supplied. John Burkett estimated the labor supply of both nursing assistants and nurses to be price elastic, with that of nursing assistants to be 1.9 (very close to that of child-care workers) and of nurses to be 1.1. Note that the price elasticity of labor supply of the higher-paid nurses is a bit lower than that of lower-paid nursing assistants.

In contrast, John Rizzo and David Blumenthal estimated the price elasticity of labor supply for young physicians (under the age of 40) to be about 0.3. This means that a 10% increase in wages leads to an increase in the quantity of labor supplied of only about 3%. In addition, when Rizzo and Blumenthal looked at labor supply elasticities by gender, they found the female physicians' labor supply price elasticity to be a bit higher (at about 0.5) than that of the males (at about 0.2) in the sample. Because earnings of female physicians in the sample were lower than earnings of the male physicians in the sample, this difference in labor supply elasticities was expected. Moreover, since the sample consisted of physicians in the early phases of their careers, the positive, though small, price elasticities were also expected. Many of the individuals in the sample also had high debt levels, often from educational loans. Thus, the chance to earn more by working more is an opportunity to repay educational and other loans.

In another study of physicians' labor supply that was not restricted to young physicians, Douglas M. Brown found the labor supply price elasticity for primary care physicians to be close to zero and that of specialists to be negative, at about −0.3. Thus, for this sample of physicians, increases in wages have little or no effect on the amount the primary care doctors work, while a 10% increase in wages for specialists *reduces* their quantity of labor supplied by about 3%. Because the earnings of specialists exceed those of primary care doctors, this elasticity differential also makes sense.

Sources: David M. Blau, "The Supply of Child Care Labor," Journal of Labor Economics 11:2 (April 1993): 324–347; David M. Brown, "The Rising Cost of Physician's Services: A Correction and Extension on Supply," Review of Economics and Statistics 76 (2) (May 1994): 389–393; John P. Burkett, "The Labor Supply of Nurses and Nursing Assistants in the United States," Eastern Economic Journal 31(4) (Fall 2005): 585–599; John A. Rizzo and Paul Blumenthal. "Physician Labor Supply: Do Income Effects Matter?" Journal of Health Economics 13:4 (December 1994): 433–453.

ANSWER TO TRY IT! PROBLEM

While at a point in time the supply of people with degrees in computer science is very price inelastic, over time the elasticity should rise. That more students were majoring in computer science lends credence to this prediction. As supply becomes more price elastic, salaries in this field should rise more slowly.

4. REVIEW AND PRACTICE

Summary

This chapter introduced a new tool: the concept of elasticity. Elasticity is a measure of the degree to which a dependent variable responds to a change in an independent variable. It is the percentage change in the dependent variable divided by the percentage change in the independent variable, all other things unchanged.

The most widely used elasticity measure is the price elasticity of demand, which reflects the responsiveness of quantity demanded to changes in price. Demand is said to be price elastic if the absolute value of the price elasticity of demand is greater than 1, unit price elastic if it is equal to 1, and price inelastic if it is less than 1. The price elasticity of demand is useful in forecasting the response of quantity demanded to price changes; it is also useful for predicting the impact a price change will have on total revenue. Total revenue moves in the direction of the quantity change if demand is price elastic, it moves in the direction of the price change if demand is price inelastic, and it does not change if demand is unit price elastic. The most important determinants of the price elasticity of demand are the availability of substitutes, the importance of the item in household budgets, and time.

Two other elasticity measures commonly used in conjunction with demand are income elasticity and cross price elasticity. The signs of these elasticity measures play important roles. A positive income elasticity tells us that a good is normal; a negative income elasticity tells us the good is inferior. A positive cross price elasticity tells us that two goods are substitutes; a negative cross price elasticity tells us they are complements.

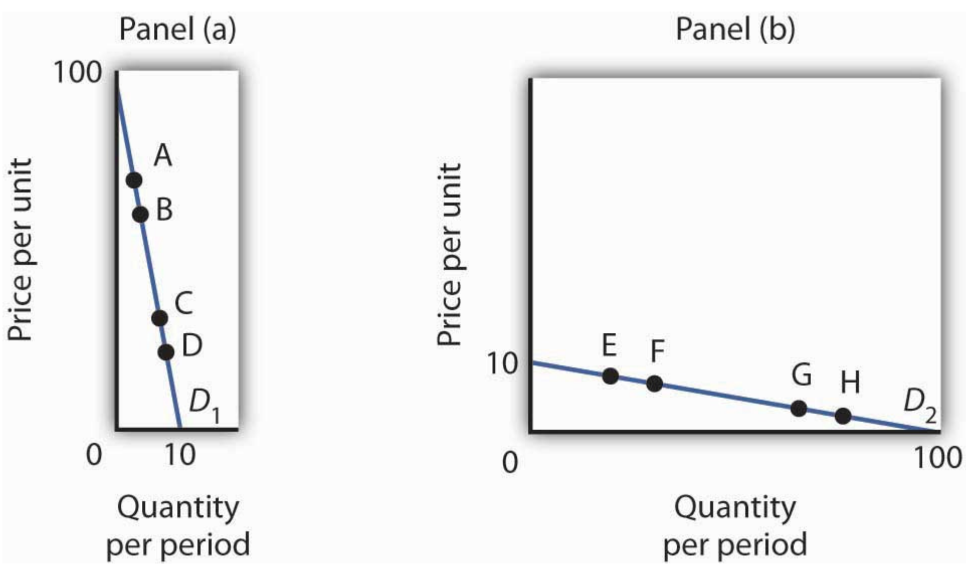
Elasticity of supply measures the responsiveness of quantity supplied to changes in price. The value of price elasticity of supply is generally positive. Supply is classified as being price elastic, unit price elastic, or price inelastic if price elasticity is greater than 1, equal to 1, or less than 1, respectively. The length of time over which supply is being considered is an important determinant of the price elasticity of supply.

CONCEPT PROBLEMS

1. Explain why the price elasticity of demand is generally a negative number, except in the cases where the demand curve is perfectly elastic or perfectly inelastic. What would be implied by a positive price elasticity of demand?
2. Explain why the sign (positive or negative) of the cross price elasticity of demand is important.
3. Explain why the sign (positive or negative) of the income elasticity of demand is important.
4. Economists Dale Heien and Cathy Roheim Wessells found that the price elasticity of demand for fresh milk is −0.63 and the price elasticity of demand for cottage cheese is −1.1.[6] Why do you think the elasticity estimates differ?
5. The price elasticity of demand for health care has been estimated to be −0.2. Characterize this demand as price elastic, unit price elastic, or price inelastic. The text argues that the greater the importance of an item in consumer budgets, the greater its elasticity. Health-care costs account for a relatively large share of household budgets. How could the price elasticity of demand for health care be such a small number?
6. Suppose you are able to organize an alliance that includes all farmers. They agree to follow the group's instructions with respect to the quantity of agricultural products they produce. What might the group seek to do? Why?
7. Suppose you are the chief executive officer of a firm, and you have been planning to reduce your prices. Your marketing manager reports that the price elasticity of demand for your product is −0.65. How will this news affect your plans?
8. Suppose the income elasticity of the demand for beans is −0.8. Interpret this number.
9. Transportation economists generally agree that the cross price elasticity of demand for automobile use with respect to the price of bus fares is about 0. Explain what this number means.
10. Suppose the price elasticity of supply of tomatoes as measured on a given day in July is 0. Interpret this number.
11. The price elasticity of supply for child-care workers was reported to be quite high, about 2. What will happen to the wages of child-care workers as demand for them increases, compared to what would happen if the measured price elasticity of supply were lower?
12. The Case in Point on cigarette taxes and teen smoking suggests that a higher tax on cigarettes would reduce teen smoking and premature deaths. Should cigarette taxes therefore be raised?

NUMERICAL PROBLEMS

1. Economist David Romer found that in introductory economics classes a 10% increase in class attendance is associated with a 4% increase in course grade.[7] What is the elasticity of course grade with respect to class attendance?
2. Refer to Figure 5.2 and
   a. Using the arc elasticity of demand formula, compute the price elasticity of demand between points B and C.
   b. Using the arc elasticity of demand formula, compute the price elasticity of demand between points D and E.
   c. How do the values of price elasticity of demand compare? Why are they the same or different?
   d. Compute the slope of the demand curve between points B and C.
   e. Computer the slope of the demand curve between points D and E.
   f. How do the slopes compare? Why are they the same or different?
3. Consider the following quote from *The Wall Street Journal:* "A bumper crop of oranges in Florida last year drove down orange prices. As juice marketers' costs fell, they cut prices by as much as 15%. That was enough to tempt some value-oriented customers: unit volume of frozen juices actually rose about 6% during the quarter."
   a. Given these numbers, and assuming there were no changes in demand shifters for frozen orange juice, what was the price elasticity of demand for frozen orange juice?
   b. What do you think happened to total spending on frozen orange juice? Why?
4. Suppose you are the manager of a restaurant that serves an average of 400 meals per day at an average price per meal of $20. On the basis of a survey, you have determined that reducing the price of an average meal to $18 would increase the quantity demanded to 450 per day.
   a. Compute the price elasticity of demand between these two points.
   b. Would you expect total revenues to rise or fall? Explain.
   c. Suppose you have reduced the average price of a meal to $18 and are considering a further reduction to $16. Another survey shows that the quantity demanded of meals will increase from 450 to 500 per day. Compute the price elasticity of demand between these two points.
   d. Would you expect total revenue to rise or fall as a result of this second price reduction? Explain.
   e. Compute total revenue at the three meal prices. Do these totals confirm your answers in (b) and (d) above?
5. The text notes that, for any linear demand curve, demand is price elastic in the upper half and price inelastic in the lower half. Consider the following demand curves:



The table gives the prices and quantities corresponding to each of the points shown on the two demand curves.

Demand curve D_1 [Panel (a)]			Demand curve D_2 [Panel (b)]		
	Price	Quantity		Price	Quantity
A	80	2	E	8	20
B	70	3	F	7	30
C	30	7	G	3	70
D	20	8	H	2	80

a. Compute the price elasticity of demand between points A and B and between points C and D on demand curve D_1 in Panel (a). Are your results consistent with the notion that a linear demand curve is price elastic in its upper half and price inelastic in its lower half?
b. Compute the price elasticity of demand between points E and F and between points G and H on demand curve D_2 in Panel (b). Are your results consistent with the notion that a linear demand curve is price elastic in its upper half and price inelastic in its lower half?
c. Compare total spending at points A and B on D_1 in Panel (a). Is your result consistent with your finding about the price elasticity of demand between those two points?
d. Compare total spending at points C and D on D_1 in Panel (a). Is your result consistent with your finding about the price elasticity of demand between those two points?
e. Compare total spending at points E and F on D_2 in Panel (b). Is your result consistent with your finding about the price elasticity of demand between those two points?
f. Compare total spending at points G and H on D_2 in Panel (b). Is your result consistent with your finding about the price elasticity of demand between those two points?

6. Suppose Janice buys the following amounts of various food items depending on her weekly income:

Weekly Income	Hamburgers	Pizza	Ice Cream Sundaes
$500	3	3	2
$750	4	2	2

a. Compute Janice's income elasticity of demand for hamburgers.
b. Compute Janice's income elasticity of demand for pizza.
c. Compute Janice's income elasticity of demand for ice cream sundaes.
d. Classify each good as normal or inferior.

7. Suppose the following table describes Jocelyn's weekly snack purchases, which vary depending on the price of a bag of chips:

Price of bag of chips	Bags of chips	Containers of salsa	Bags of pretzels	Cans of soda
$1.00	2	3	1	4
$1.50	1	2	2	4

a. Compute the cross price elasticity of salsa with respect to the price of a bag of chips.
b. Compute the cross price elasticity of pretzels with respect to the price of a bag of chips.
c. Compute the cross price elasticity of soda with respect to the price of a bag of chips.
d. Are chips and salsa substitutes or complements? How do you know?
e. Are chips and pretzels substitutes or complements? How do you know?
f. Are chips and soda substitutes or complements? How do you know?

8. The table below describes the supply curve for light bulbs:

Price per light bulb	Quantity supplied per day
$1.00	500
1.50	3,000
2.00	4,000
2.50	4,500
3.00	4,500

Compute the price elasticity of supply and determine whether supply is price elastic, price inelastic, perfectly elastic, perfectly inelastic, or unit elastic:

a. when the price of a light bulb increases from \$1.00 to \$1.50.
b. when the price of a light bulb increases from \$1.50 to \$2.00.
c. when the price of a light bulb increases from \$2.00 to \$2.50.
d. when the price of a light bulb increases from \$2.50 to \$3.00.

ENDNOTES

1. Notice that since the number of units sold of a good is the same as the number of units bought, the definition for total revenue could also be used to define total spending. Which term we use depends on the question at hand. If we are trying to determine what happens to revenues of sellers, then we are asking about total revenue. If we are trying to determine how much consumers spend, then we are asking about total spending.
2. Division by zero results in an undefined solution. Saying that the price elasticity of demand is infinite requires that we say the denominator "approaches" zero.
3. See Georges Bresson, Joyce Dargay, Jean-Loup Madre, and Alain Pirotte, "Economic and Structural Determinants of the Demand for French Transport: An Analysis on a Panel of French Urban Areas Using Shrinkage Estimators." *Transportation Research: Part A* 38: 4 (May 2004): 269–285. See also Anna Matas. "Demand and Revenue Implications of an Integrated Transport Policy: The Case of Madrid." *Transport Reviews* 24:2 (March 2004): 195–217.
4. Robert B. Ekelund, S. Ford, and John D. Jackson. "Are Local TV Markets Separate Markets?" *International Journal of the Economics of Business* 7:1 (2000): 79–97.
5. Although close to zero in all cases, the significant and positive signs of income elasticity for marijuana, alcohol, and cocaine suggest that they are normal goods, but significant and negative signs, in the case of heroin, suggest that heroin is an inferior good; Saffer and Chaloupka (cited below) suggest the effects of income for all four substances might be affected by education.

   Sources: John A. Tauras. "Public Policy and Smoking Cessation among Young Adults in the United States," *Health Policy*, 68:3 (June 2004): 321–332. Georges Bresson, Joyce Dargay, Jean-Loup Madre, and Alain Pirotte, "Economic and Structural Determinants of the Demand for French Transport: An Analysis on a Panel of French Urban Areas Using Shrinkage Estimators," *Transportation Research: Part A* 38:4 (May 2004): 269–285; Avner Bar-Ilan and Bruce Sacerdote, "The Response of Criminals and Non-Criminals to Fines," *Journal of Law and Economics*, 47:1 (April 2004): 1–17; Hana Ross and Frank J. Chaloupka, "The Effect of Public Policies and Prices on Youth Smoking," *Southern Economic Journal* 70:4 (April 2004): 796–815; Anna Matas, "Demand and Revenue Implications of an Integrated Transport Policy: The Case of Madrid," *Transport Reviews*, 24:2 (March 2004): 195–217; Matthew C. Farrelly, Terry F. Pechacek, and Frank J. Chaloupka; "The Impact of Tobacco Control Program Expenditures on Aggregate Cigarette Sales: 1981–2000," *Journal of Health Economics* 22:5 (September 2003): 843–859; Robert B. Ekelund, S. Ford, and John D. Jackson. "Are Local TV Markets Separate Markets?" *International Journal of the Economics of Business* 7:1 (2000): 79–97; Henry Saffer and Frank Chaloupka, "The Demand for Illicit Drugs," *Economic Inquiry* 37(3) (July, 1999): 401–411; Robert W. Fogel, "Catching Up With the Economy," *American Economic Review* 89(1) (March, 1999):1–21; Michael Grossman, "A Survey of Economic Models of Addictive Behavior," *Journal of Drug Issues* 28:3 (Summer 1998):631–643; Sanjib Bhuyan and Rigoberto A. Lopez, "Oligopoly Power in the Food and Tobacco Industries," *American Journal of Agricultural Economics* 79 (August 1997):1035–1043; Michael Grossman, "Cigarette Taxes," *Public Health Reports* 112:4 (July/August 1997): 290–297; Ann Hansen, "The Tax Incidence of the Colorado State Lottery Instant Game," *Public Finance Quarterly* 23(3) (July, 1995):385–398; Daniel B. Suits, "Agriculture," in Walter Adams and James Brock, eds., *The Structure of American Industry*, 9th ed. (Englewood Cliffs: Prentice Hall, , 1995), pp. 1–33; Kenneth G. Elzinga, "Beer," in Walter Adams and James Brock, eds., *The Structure of American Industry*, 9th ed. (Englewood Cliffs: Prentice Hall, 1995), pp. 119–151; John A. Rizzo and David Blumenthal, "Physician Labor Supply: Do Income Effects Matter?" *Journal of Health Economics* 13(4) (December 1994):433–453; Douglas M. Brown, "The Rising Price of Physicians' Services: A Correction and Extension on Supply," *Review of Economics and Statistics* 76(2) (May 1994):389–393; George C. Davis and Michael K. Wohlgenant, "Demand Elasticities from a Discrete Choice Model: The Natural Christmas Tree Market," *Journal of Agricultural Economics* 75(3) (August 1993):730–738; David M. Blau, "The Supply of Child Care Labor," *Journal of Labor Economics* 2(11) (April 1993):324–347; Richard Blundell *et al*., "What Do We Learn About Consumer Demand Patterns from Micro Data?", *American Economic Review* 83(3) (June 1993):570–597; F. Gasmi, *et al*., "Econometric Analysis of Collusive Behavior in a Soft-Drink Market," *Journal of Economics and Management Strategy* (Summer 1992), pp. 277–311; M.R. Baye, D.W. Jansen, and J.W. Lee, "Advertising Effects in Complete Demand Systems," *Applied Economics* 24 (1992):1087–1096; Gary W. Brester and Michael K. Wohlgenant, "Estimating Interrelated Demands for Meats Using New Measures for Ground and Table Cut Beef," *American Journal of Agricultural Economics* 73 (November 1991):1182–1194; Adesoji, O. Adelaja, "Price Changes, Supply Elasticities, Industry Organization, and Dairy Output Distribution," *American Journal of Agricultural Economics* 73:1 (February 1991):89–102; Mark A. R. Kleinman, *Marijuana: Costs of Abuse, Costs of Control* (NY:Greenwood Press, 1989); Jules M. Levine, *et al*., "The Demand for Higher Education in Three Mid-Atlantic States," *New York Economic Review* 18 (Fall 1988):3–20; Dale Heien and Cathy Roheim Wessells, "The Demand for Dairy Products: Structure, Prediction, and Decomposition," *American Journal of Agriculture Economics* (May 1988):219–228; Michael Grossman and Henry Saffer, "Beer Taxes, the Legal Drinking Age, and Youth Motor Vehicle Fatalities," *Journal of Legal Studies* 16(2) (June 1987):351–374; James M. Griffin and Henry B. Steele, *Energy Economics and Policy* (New York: Academic Press, 1980), p. 232.
6. Dale M. Heien and Cathy Roheim Wessels, "The Demand for Dairy Products: Structure, Prediction, and Decomposition," *American Journal of Agricultural Economics* 70:2 (May 1988): 219–228.
7. David Romer, "Do Students Go to Class? Should They?" *Journal of Economic Perspectives* 7:3 (Summer 1993): 167–174.

CHAPTER 6

Markets, Maximizers, and Efficiency

START UP: A DRIVE IN THE COUNTRY

Suppose you decide to take a drive. For purposes of this example, we will assume that you have a car available, that the weather is pleasant, and that there is an area nearby that will be perfect for your drive.

Your decision to take this drive is a choice. Since economics deals with choices, we can put economics to work in thinking about it. Economists assume that people make choices that maximize the value of some objective. You are a consumer; we assume that taking a drive is a choice that maximizes your utility—the satisfaction you obtain from your use of goods and services and from the activities you pursue.

You certainly plan to enjoy the drive; that enjoyment is the benefit you expect from it. But you will give up some things as well. Your drive will take some time, time you could have spent doing something else. It will take some gasoline; what you spend for the gasoline could have been used for something else. The drive will also generate some wear and tear on your car. That will cost you the price of repair and maintenance and reduced resale value of your car. The opportunity cost of your drive will thus include the value of the best other use of your time and the value of the best other use of the funds your drive will require. To maximize utility you will weigh the benefits of the drive against the cost of the drive and maximize the difference between those benefits and costs.

This chapter introduces the method through which maximizing choices can be made. This method applies not just to your decision to take a drive, but also to Wal-Mart's decision to hire extra workers and to USX Corporation's to produce extra steel. The method we will learn can be applied to the analysis of any choice; we will use it throughout our investigation of microeconomics.

We will also see how maximizing choices by individuals and by firms can lead to an allocation of resources that generates the greatest gains possible for the economy as a whole. In this analysis, we will put a new item in our toolkit, the method through which individuals and firms maximize, together with demand and supply analysis, to see how the marketplace can guide resources to their best uses.

We will also examine cases in which maximizing choices do not guide resources to their best uses. That possibility is suggested by another aspect of your choice to take a drive. In addition to the costs you will consider, there will be costs imposed on others. Your drive will pollute the air, so part of the opportunity cost of the drive will be the value of the slightly cleaner air people in your area might have had. Resources such as the air we breathe will almost certainly be misallocated as the result of maximizing choices. We will see just how misallocation of an economy's resources can occur and how this misallocation could be fixed.

1. THE LOGIC OF MAXIMIZING BEHAVIOR

LEARNING OBJECTIVES

1. **Explain the maximization assumption that economists make in explaining the behavior of consumers and firms.**
2. **Explain and illustrate the concepts of marginal benefit and marginal cost and apply them to understanding the marginal decision rule.**

economic profit

The difference between total revenue and total cost.

To say that individuals maximize is to say that they pick some objective and then seek to maximize its value. A sprinter might want to maximize his or her speed; a politician might want to maximize the probability that he or she will win the next election. Economists pay special attention to two groups of maximizers: consumers and firms. We assume that consumers seek to maximize utility and that firms seek to maximize **economic profit**, which is the difference between total revenue and total cost. The costs involved in this concept of economic profit are computed in the economic sense—as the opportunity costs, or value of the best opportunity forgone.

The assumption of maximizing behavior lies at the heart of economic analysis. As we explore its implications, however, we must keep in mind the distinction between models and the real world. Our model assumes that individuals make choices in a way that achieves a maximum value for some clearly defined objective. In using such a model, economists do not assume that people actually go through the calculations we will describe. What economists do argue is that people's behavior is broadly consistent with such a model. People may not consciously seek to maximize anything, but they behave as though they do.

1.1 The Analysis of Maximizing Behavior

net benefit

The total benefit of an activity minus its opportunity cost.

The activities of consumers and firms have benefits, and they also have opportunity costs. We assume that given these benefits and costs, consumers and firms will make choices that maximize the **net benefit** of each activity—the total benefit of the activity minus its opportunity cost. The specific measures of benefit and cost vary with the kind of choice being made. In the case of a firm's choices in production, for example, the total benefit of production is the revenue a firm receives from selling the product; the total cost is the opportunity cost the firm incurs by producing it. The net benefit is thus total revenue minus total opportunity cost, or economic profit.

Economists maintain that in order to maximize net benefit, consumers and firms evaluate each activity at the margin—they consider the additional benefit and the additional cost of another unit of the activity. Should you "supersize" your order at McDonald's? Will the additional beverage and the additional french fries be worth the extra cost? Should a firm hire one more worker? Will the benefits to the firm of hiring this worker be worth the additional cost of hiring him or her?

marginal benefit

The amount by which an additional unit of an activity increases its total benefit.

marginal cost

The amount by which an additional unit of an activity increases its total cost.

The **marginal benefit** is the amount by which an additional unit of an activity increases its total benefit. It is the amount by which the extra french fries increase your satisfaction, or the extra revenue the firm expects to bring in by hiring another worker. The **marginal cost** is the amount by which an additional unit of an activity increases its total cost. You will pay more to supersize your McDonald's order; the firm's labor costs will rise when it hires another worker.

marginal decision rule

If the marginal benefit of an additional unit of an activity exceeds the marginal cost, the quantity of the activity should be increased. If the marginal benefit is less than the marginal cost, the quantity should be reduced.

To determine the quantity of any activity that will maximize its net benefit, we apply the **marginal decision rule**: If the marginal benefit of an additional unit of an activity exceeds the marginal cost, the quantity of the activity should be increased. If the marginal benefit is less than the marginal cost, the quantity should be reduced. Net benefit is maximized at the point at which marginal benefit equals marginal cost. The marginal decision rule is at the heart of the economic way of thinking. The rule basically says this: If the additional benefit of one more unit exceeds the extra cost, do it; if not, do not. This simple logic gives us a powerful tool for the analysis of choice. Perhaps more than any other rule in economic analysis, the marginal decision rule typifies the way in which economists analyze problems. We shall apply it in every chapter that follows in the microeconomics portion of this text.

Maximizing choices must be made within the parameters imposed by some **constraint**, which is a boundary that limits the range of choices that can be made. We assume that a consumer seeks the greatest satisfaction possible within the limits of his or her income or budget. A firm cannot produce beyond the limits of its production capacity at a point in time.

constraint

A boundary that limits the range of choices that can be made.

The marginal decision rule forms the foundation for the structure economists use to analyze all choices. At first glance, it may seem that a consumer seeking satisfaction from, say, pizza has little in common with an entrepreneur seeking profit from the production of custom-designed semiconductors. But maximizing choices always follow the marginal decision rule—and that rule holds regardless of what is being maximized or who is doing the maximizing.

To see how the logic of maximizing choices works, we will examine a specific problem. We will then extend that problem to the general analysis of maximizing choices.

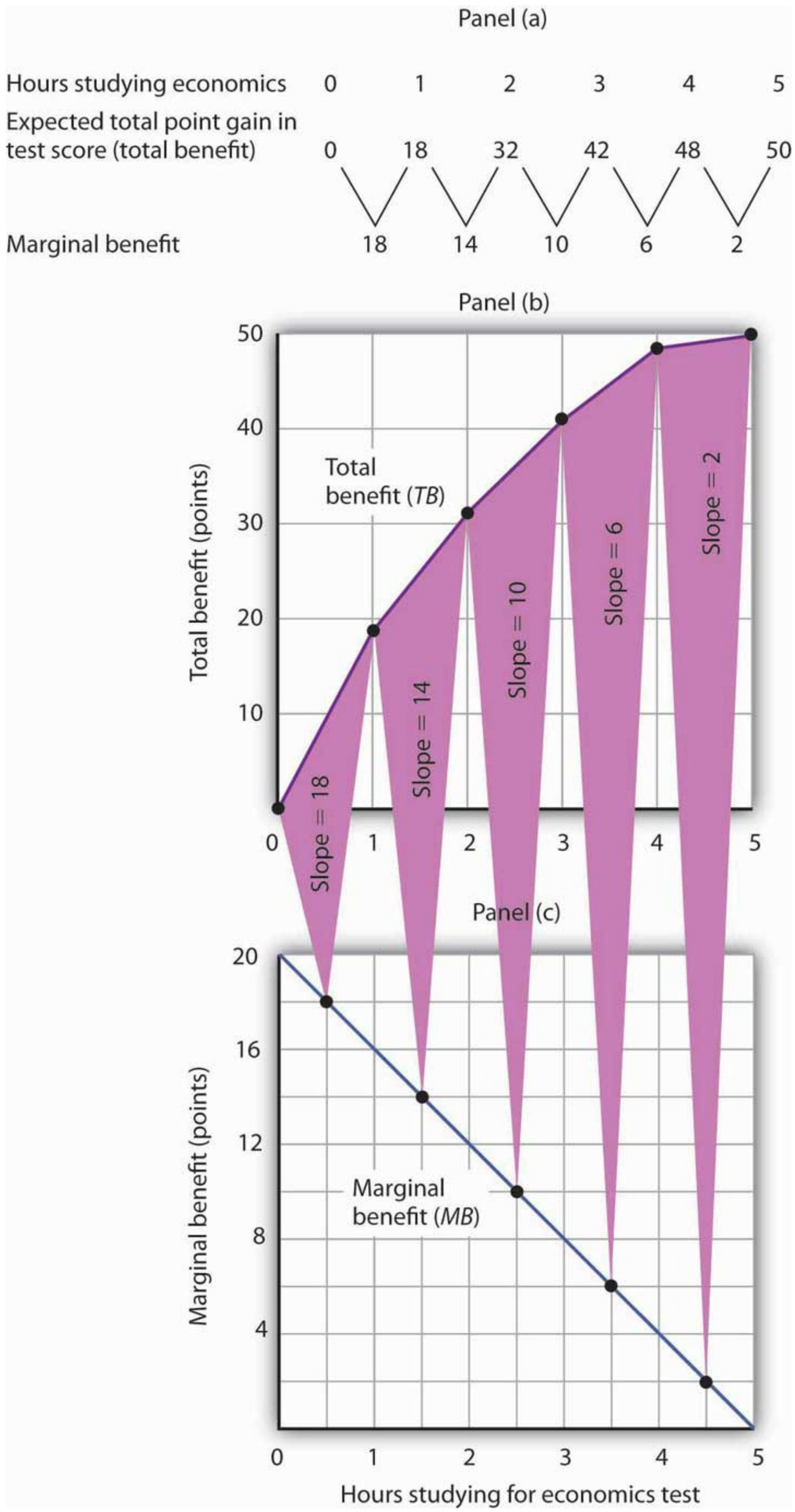
A Problem in Maximization

Suppose a college student, Laurie Phan, faces two midterms tomorrow, one in economics and another in accounting. She has already decided to spend 5 hours studying for the two examinations. This decision imposes a constraint on the problem. Suppose that Ms. Phan's goal is to allocate her 5 hours of study so that she increases her total score for the two exams by as much as possible.

Ms. Phan expects the relationship between the time she spends studying for the economics exam and the total gain in her score to be as given by the second row of the table in Panel (a) of Figure 6.1. We interpret the expected total gain in her score as the total benefit of study. She expects that 1 hour of study will raise her score by 18 points; 2 hours will raise it by 32 points, and so on. These values are plotted in Panel (b). Notice that the total benefit curve rises, but by smaller and smaller amounts, as she studies more and more. The slope of the curve, which in this case tells us the rate at which her expected score rises with increased study time, falls as we travel up and to the right along the curve.

FIGURE 6.1 The Benefits of Studying Economics

The table in Panel (a) shows the total benefit and marginal benefit of the time Laurie Phan spends studying for her economics exam. Panel (b) shows the total benefit curve. Panel (c) shows the marginal benefit curve, which is given by the slope of the total benefit curve in Panel (b).



Now look at the third row in the table in Panel (a). It tells us the amount by which each additional hour of study increases her expected score; it gives the marginal benefit of studying for the economics exam. Marginal benefit equals the amount by which total benefit rises with each additional hour of study. Because these marginal benefits are given by the changes in total benefits from additional hours of study, they equal the slope of the total benefit curve. We see this in the relationship between Panels (b) and (c) of Figure 6.1. The decreasing slope of the total benefit curve in Panel (b) gives us the downward-sloping marginal benefit curve in Panel (c).

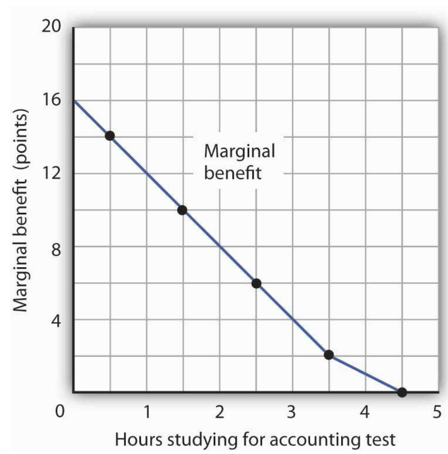
The marginal benefit curve tells us what happens when we pass from one point to another on the total benefit curve, so we have plotted marginal benefits at the midpoints of the hourly intervals in Panel (c). For example, the total benefit curve in Panel (b) tells us that, when Ms. Phan increases her time studying for the economics exam from 2 hours to 3 hours, her total benefit rises from 32 points to 42 points. The increase of 10 points is the marginal benefit of increasing study time for the economics exam from 2 hours to 3 hours. We mark the point for a marginal benefit of 10 points midway between 2 and 3 hours. Because marginal values tell us what happens as we pass from one quantity to the next, we shall always plot them at the midpoints of intervals of the variable on the horizontal axis.

We can perform the same kind of analysis to obtain the marginal benefit curve for studying for the accounting exam. Figure 6.2 presents this curve. Like the marginal benefit curve for studying economics, it slopes downward. Once again, we have plotted marginal values at the midpoints of the intervals. Increasing study time in accounting from 0 to 1 hour increases Ms. Phan's expected accounting score by 14 points.

Ms. Phan's marginal benefit curves for studying typify a general phenomenon in economics. Marginal benefit curves for virtually all activities, including the activities of consumers and of firms, slope downward. Think about your own experience with studying. On a given day, the first hour spent studying a certain subject probably generates a greater marginal benefit than the second, and the second hour probably generates a greater marginal benefit than the third. You may reach a point at which an extra hour of study is unlikely to yield any benefit at all. Of course, our example of Laurie Phan's expected exam scores is a highly stylized one. One could hardly expect a student to have a precise set of numbers to guide him or her in allocating study time. But it is certainly the case that students have a rough idea of the likely payoff of study time in different subjects. If you were faced with exams in two subjects, it is likely that you would set aside a certain amount of study time, just as Ms. Phan did in our example. And it is likely that your own experience would serve as a guide in determining how to allocate that time. Economists do not assume that people have numerical scales in their heads with which to draw marginal benefit and marginal cost curves. They merely assume that people act as if they did.

FIGURE 6.2 The Marginal Benefits of Studying Accounting

The marginal benefit Laurie Phan expects from studying for her accounting exam is shown by the marginal benefit curve. The first hour of study increases her expected score by 14 points, the second hour by 10 points, the third by 6 points, and so on.



The nature of marginal benefits can change with different applications. For a restaurant, the marginal benefit of serving one more meal can be defined as the revenue that meal produces. For a consumer, the marginal benefit of one more slice of pizza can be considered in terms of the additional satisfaction the pizza will create. But whatever the nature of the benefit, marginal benefits generally fall as quantities increase.

Ms. Phan's falling marginal benefit from hours spent studying accounting has special significance for our analysis of her choice concerning how many hours to devote to economics. In our problem, she had decided to devote 5 hours to studying the two subjects. That means that the opportunity cost of an hour spent studying economics equals the benefit she would have gotten spending that hour studying accounting.

Suppose, for example, that she were to consider spending all 5 hours studying accounting. The marginal benefit curve for studying for her accounting exam tells us that she expects that the fifth hour will add nothing to her score. Shifting that hour to economics would cost nothing. We can say that the marginal cost of the first hour spent studying economics is zero. We obtained this value from the marginal benefit curve for studying accounting in Figure 6.2.

Similarly, we can find the marginal cost of the second hour studying economics. That requires giving up the fourth hour spent on accounting. Figure 6.2 tells us that the marginal benefit of that hour equals 2—that is the marginal cost of spending the second hour studying economics.

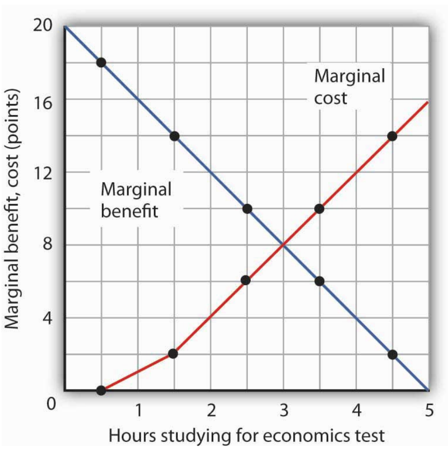
Figure 6.3 shows the marginal cost curve of studying economics. We see that at first, time devoted to studying economics has a low marginal cost. As time spent studying economics increases, however, it requires her to give up study time in accounting that she expects will be more and more productive. The marginal cost curve for studying economics can thus be derived from the marginal benefit curve for studying accounting. Figure 6.3 also shows the marginal benefit curve for studying economics that we derived in Panel (b) of Figure 6.1.

Just as marginal benefit curves generally slope downward, marginal cost curves generally slope upward, as does the one in Figure 6.3. In the case of allocating time, the phenomenon of rising marginal cost results from the simple fact that, the more time a person devotes to one activity, the less time is available for another. And the more one reduces the second activity, the greater the forgone marginal benefits are likely to be. That means the marginal cost curve for that first activity rises.

Because we now have marginal benefit and marginal cost curves for studying economics, we can apply the marginal decision rule. This rule says that, to maximize the net benefit of an activity, a decision maker should increase an activity up to the point at which marginal benefit equals marginal cost. That occurs where the marginal benefit and marginal cost curves intersect, with 3 hours spent studying economics and 2 hours spent studying accounting.

FIGURE 6.3 **The Marginal Benefits and Marginal Costs of Studying Economics**

The marginal benefit curve from Panel (c) of Figure 6.1 is shown together with the marginal costs of studying economics. The marginal cost curve is derived from the marginal benefit curve for studying accounting shown in Figure 6.2.



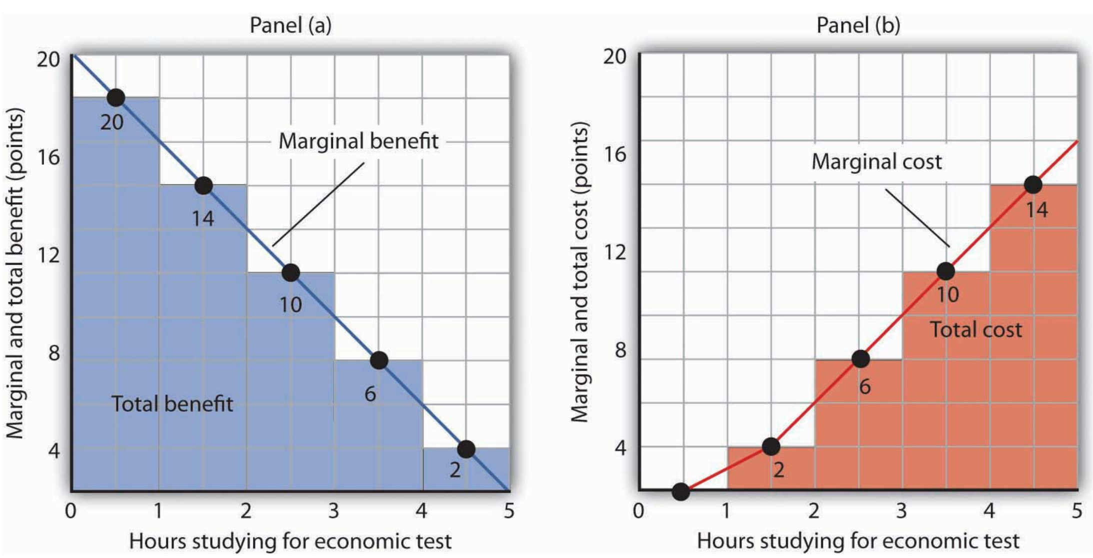
Using Marginal Benefit and Marginal Cost Curves to Find Net Benefits

We can use marginal benefit and marginal cost curves to show the total benefit, the total cost, and the net benefit of an activity. We will see that equating marginal benefit to marginal cost does, indeed, maximize net benefit. We will also develop another tool to use in interpreting marginal benefit and cost curves.

Panel (a) of Figure 6.4 shows the marginal benefit curve we derived in Panel (c) of Figure 6.1. The corresponding point on the marginal benefit curve gives the marginal benefit of the first hour of study for the economics exam, 18 points. This same value equals the area of the rectangle bounded by 0 and 1 hour of study and the marginal benefit of 18. Similarly, the marginal benefit of the second hour, 14 points, is shown by the corresponding point on the marginal benefit curve and by the area of the shaded rectangle bounded by 1 and 2 hours of study. The total benefit of 2 hours of study equals the sum of the areas of the first two rectangles, 32 points. We continue this procedure through the fifth hour of studying economics; the areas for each of the shaded rectangles are shown in the graph.

FIGURE 6.4 **The Benefits and Costs of Studying Economics**

Panel (a) shows the marginal benefit curve of Figure 6.1. The total benefit of studying economics at any given quantity of study time is given approximately by the shaded area below the marginal benefit curve up to that level of study. Panel (b) shows the marginal cost curve from Figure 6.3. The total cost of studying economics at any given quantity of study is given approximately by the shaded area below the marginal cost curve up to that level of study.



Two features of the curve in Panel (a) of Figure 6.4 are particularly important. First, note that the sum of the areas of the five rectangles, 50 points, equals the total benefit of 5 hours of study given in the table in Panel (a) of Figure 6.1. Second, notice that the shaded areas are approximately equal to the area under the marginal benefit curve between 0 and 5 hours of study. We can pick any quantity of study time, and the total benefit of that quantity equals the sum of the shaded rectangles between zero and that quantity. Thus, the total benefit of 2 hours of study equals 32 points, the sum of the areas of the first two rectangles.

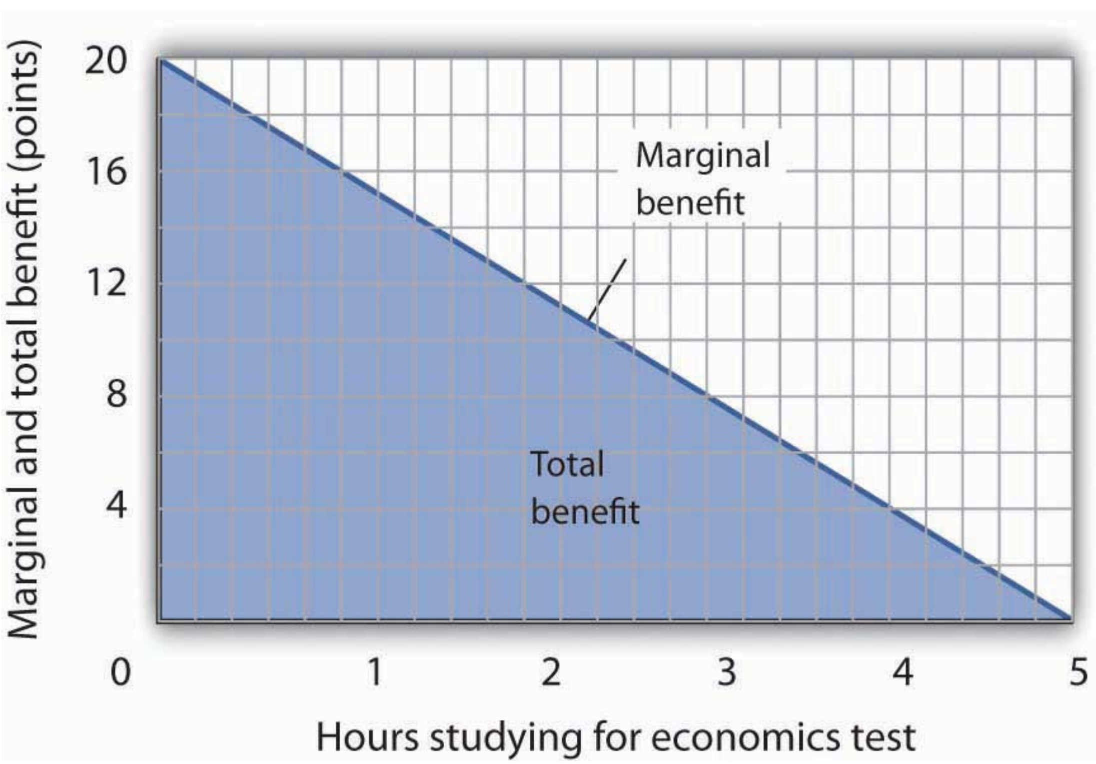
Now consider the marginal cost curve in Panel (b) of Figure 6.4. The areas of the shaded rectangles equal the values of marginal cost. The marginal cost of the first hour of study equals zero; there is thus no rectangle under the curve. The marginal cost of the second hour of study equals 2 points; that is the area of the rectangle bounded by 1 and 2 hours of study and a marginal cost of 2. The marginal cost of the third hour of study is 6 points; this is the area of the shaded rectangle bounded by 2 and 3 hours of study and a marginal cost of 6.

Looking at the rectangles in Panel (b) over the range of 0 to 5 hours of study, we see that the areas of the five rectangles total 32, the total cost of spending all 5 hours studying economics. And looking at the rectangles, we see that their area is approximately equal to the area under the marginal cost curve between 0 and 5 hours of study.

We have seen that the areas of the rectangles drawn with Laurie Phan's marginal benefit and marginal cost curves equal the total benefit and total cost of studying economics. We have also seen that these areas are roughly equal to the areas under the curves themselves. We can make this last statement much stronger. Suppose, instead of thinking in intervals of whole hours, we think in terms of smaller intervals, say, of 12 minutes. Then each rectangle would be only one-fifth as wide as the rectangles we drew in Figure 6.4. Their areas would still equal the total benefit and total cost of study, and the sum of those areas would be closer to the area under the curves. We have done this for Ms. Phan's marginal benefit curve in Figure 6.5; notice that the areas of the rectangles closely approximate the area under the curve. They still "stick out" from either side of the curve as did the rectangles we drew in Figure 6.4, but you almost need a magnifying glass to see that. The smaller the interval we choose, the closer the areas under the marginal benefit and marginal cost curves will be to total benefit and total cost. For purposes of our model, we can imagine that the intervals are as small as we like. Over a particular range of quantity, the area under a marginal benefit curve equals the total benefit of that quantity, and the area under the marginal cost curve equals the total cost of that quantity.

FIGURE 6.5 The Marginal Benefit Curve and Total Benefit

When the increments used to measure time allocated to studying economics are made smaller, in this case 12 minutes instead of whole hours, the area under the marginal benefit curve is closer to the total benefit of studying that amount of time.



Panel (a) of Figure 6.6 shows marginal benefit and marginal cost curves for studying economics, this time without numbers. We have the usual downward-sloping marginal benefit curve and upward-sloping marginal cost curve. The marginal decision rule tells us to choose D hours studying economics, the quantity at which marginal benefit equals marginal cost at point C. We know that the total benefit of study equals the area under the marginal benefit curve over the range from A to D hours of study, the area ABCD. Total cost equals the area under the marginal cost curve over the same range, or ACD. The difference between total benefit and total cost equals the area between marginal benefit and marginal

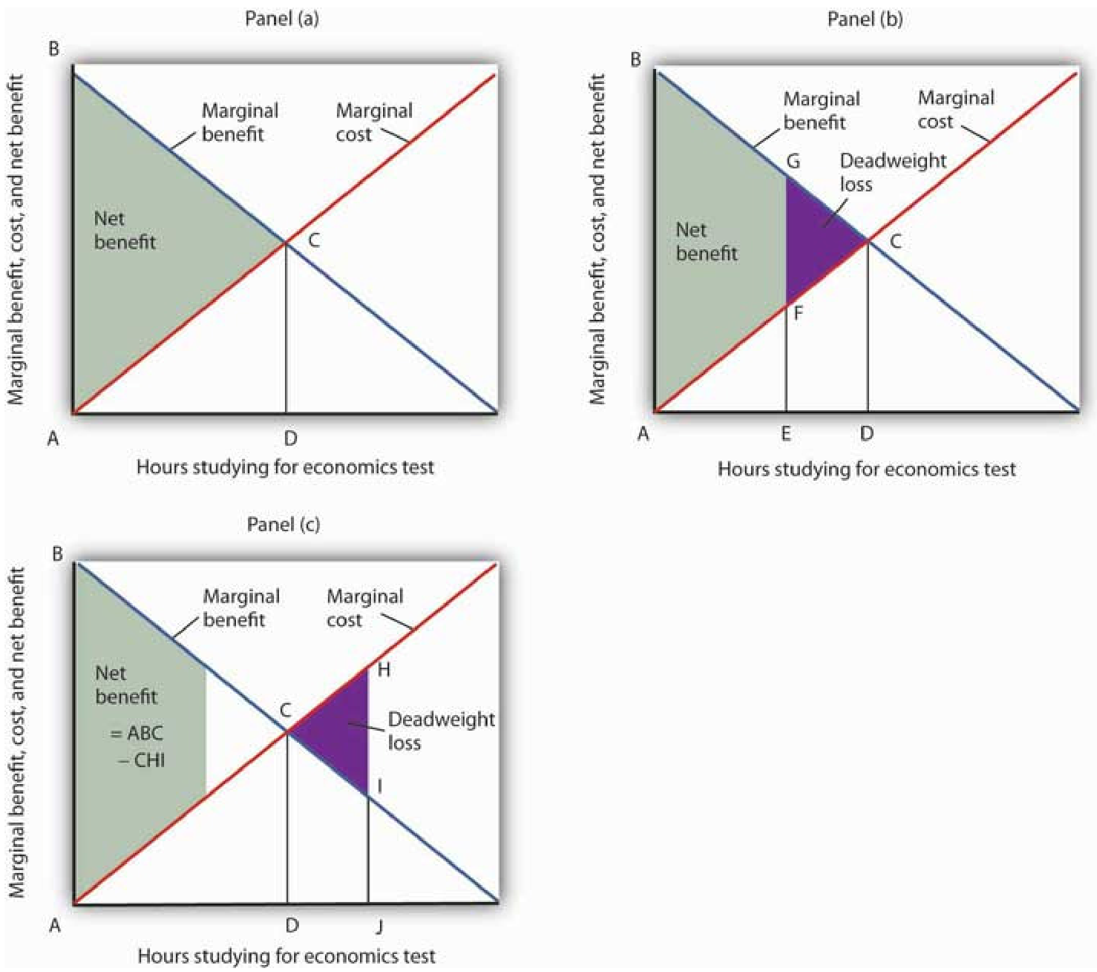
deadweight loss

The loss in net benefits resulting from a failure to carry out an activity at the most efficient level.

cost between A and D hours of study; it is the green-shaded triangle ABC. This difference is the net benefit of time spent studying economics. Panel (b) of Figure 6.6 introduces another important concept. If an activity is carried out at a level less than the efficient level, then net benefits are forgone. The loss in net benefits resulting from a failure to carry out an activity at the efficient level is called a **deadweight loss**.

FIGURE 6.6 Using Marginal Benefit and Marginal Cost Curves to Determine Net Benefit

In Panel (a) net benefits are given by the difference between total benefits (as measured by the area under the marginal benefit curve up to any given level of activity) and total costs (as measured by the area under the marginal cost curve up to any given level of activity). Maximum net benefits are found where the marginal benefit curve intersects the marginal cost curve at activity level D. Panel (b) shows that if the level of the activity is restricted to activity level E, net benefits are reduced from the light-green shaded triangle ABC in Panel (a) to the smaller area ABGF. The forgone net benefits, or deadweight loss, is given by the purple-shaded area FGC. If the activity level is increased from D to J, as shown in Panel (c), net benefits declined by the deadweight loss measured by the area CHI.



Now suppose a person increases study time from D to J hours as shown in Panel (c). The area under the marginal cost curve between D and J gives the total cost of increasing study time; it is DCHJ. The total benefit of increasing study time equals the area under the marginal benefit curve between D and J; it is DCIJ. The cost of increasing study time in economics from D hours to J hours exceeds the benefit. This gives us a deadweight loss of CHI. The net benefit of spending J hours studying economics equals the net benefit of studying for D hours less the deadweight loss, or ABC *minus* CHI. Only by studying up to the point at which marginal benefit equals marginal cost do we achieve the maximum net benefit shown in Panel (a).

We can apply the marginal decision rule to the problem in Figure 6.6 in another way. In Panel (b), a person studies economics for E hours. Reading up to the marginal benefit curve, we reach point G. Reading up to the marginal cost curve, we reach point F. Marginal benefit at G exceeds marginal cost at F; the marginal decision rule says economics study should be increased, which would take us toward the intersection of the marginal benefit and marginal cost curves. Spending J hours studying economics, as shown in Panel (c), is too much. Reading up to the marginal benefit and marginal cost curves, we see that marginal cost exceeds marginal benefit, suggesting that study time be reduced.

This completes our introduction to the marginal decision rule and the use of marginal benefit and marginal cost curves. We will spend the remainder of the chapter applying the model.

Heads Up!

It is easy to make the mistake of assuming that if an activity is carried out up to the point where marginal benefit equals marginal cost, then net benefits must be zero. Remember that following the marginal decision rule and equating marginal benefits and costs *maximizes* net benefits. It makes the *difference* between total benefits and total cost as large as possible.

KEY TAKEAWAYS

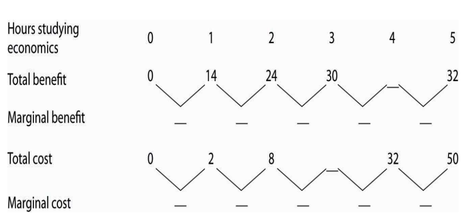
- Economists assume that decision makers make choices in the way that maximizes the value of some objective.
- Maximization involves determining the change in total benefit and the change in total cost associated with each unit of an activity. These changes are called marginal benefit and marginal cost, respectively.
- If the marginal benefit of an activity exceeds the marginal cost, the decision maker will gain by increasing the activity.
- If the marginal cost of an activity exceeds the marginal benefit, the decision maker will gain by reducing the activity.
- The area under the marginal benefit curve for an activity gives its total benefit; the area under the marginal cost curve gives the activity's total cost. Net benefit equals total benefit less total cost.
- The marginal benefit rule tells us that we can maximize the net benefit of any activity by choosing the quantity at which marginal benefit equals marginal cost. At this quantity, the net benefit of the activity is maximized.

TRY IT!

Suppose Ms. Phan still faces the exams in economics and in accounting, and she still plans to spend a total of 5 hours studying for the two exams. However, she revises her expectations about the degree to which studying economics and accounting will affect her scores on the two exams. She expects studying economics will add somewhat less to her score, and she expects studying accounting will add more. The result is the table below of expected total benefits and total costs of hours spent studying economics. Notice that several values in the table have been omitted. Fill in the missing values in the table. How many hours of study should Ms. Phan devote to economics to maximize her net benefit?

Hours studying economics	0	1	2	3	4	5
Total benefit	0	14	24	30		32
Total cost	0	2	8		32	50
Net benefit	0	12		12	0	−18

Now compute the marginal benefits and costs of hours devoted to studying economics, completing the table below.



Draw the marginal benefit and marginal cost curves for studying economics (remember to plot marginal values at the midpoints of the respective hourly intervals). Do your curves intersect at the "right" number of hours of study—the number that maximizes the net benefit of studying economics?

Case in Point: Preventing Oil Spills

Do we spill enough oil in our oceans and waterways? It is a question that perhaps only economists would ask—and, as economists, we should ask it.

There is, of course, no virtue in an oil spill. It destroys wildlife and fouls shorelines. Cleanup costs can be tremendous. However, Preventing oil Spills has costs as well: greater enforcement expenditures and higher costs to shippers of oil and, therefore, higher costs of goods such as gasoline to customers. The only way to prevent oil spills completely is to stop shipping oil. That is a cost few people would accept. But what is the right balance between environmental protection and the satisfaction of consumer demand for oil?

Vanderbilt University economist Mark Cohen examined the U.S. Coast Guard's efforts to reduce oil spills through its enforcement of shipping regulations in coastal waters and on rivers. He focused on the costs and benefits resulting from the Coast Guard's enforcement efforts in 1981. On the basis of the frequency of oil spills before the Coast Guard began its enforcement, Mr. Cohen estimated that the Coast Guard prevented 1,159,352 gallons of oil from being spilled in 1981.

Given that there was a total of 824,921 gallons of oil actually spilled in 1981, should the Coast Guard have attempted to prevent even more spillage? Mr. Cohen estimated that the marginal benefit of preventing one more gallon from being spilled was $7.27 ($3.42 in cleanup costs, $3 less in environmental damage, and $0.85 worth of oil saved). The marginal cost of preventing one more gallon from being spilled was $5.50. Mr. Cohen suggests that because the marginal benefit of more vigorous enforcement exceeded the marginal cost, more vigorous Coast Guard efforts would have been justified.

More vigorous efforts have, indeed, been pursued. In 1989, the Exxon oil tanker Exxon Valdez ran aground, spilling 10.8 million gallons of oil off the coast of Alaska. The spill damaged the shoreline of a national forest, four national wildlife refuges, three national parks, five state parks, four critical habitat areas, and a state game refuge. Exxon was ordered to pay $900 million in damages; a federal jury found Exxon and the captain guilty of criminal negligence and imposed an additional $5 billion in punitive damages. In 2008, The Supreme Court reduced the assessment of punitive damages to $507 million, with the majority arguing that the original figure was too high in comparison to the compensatory damages for a case in which the actions of the defendant, Exxon, were "reprehensible" but not intentional.

Perhaps the most important impact of the Exxon Valdez disaster was the passage of the Oil Pollution Act of 1990. It increased shipper liability from $14 million to $100 million. It also required double-hulled tankers for shipping oil.

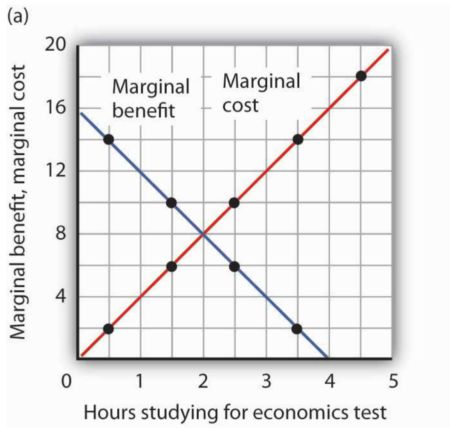
The European Union (EU) has also strengthened its standards for oil tankers. The 2002 breakup of the oil tanker Prestige off the coast of Spain resulted in the spillage of 3.2 million gallons of oil. The EU had planned to ban single-hulled tankers, phasing in the ban between 2003 and 2015. The sinking of the Prestige led the EU to move up that deadline.

Spill crises have led both the United States and the European Union to tighten up their regulations of oil tankers. The result has been a reduction in the quantity of oil spilled, which was precisely what economic research had concluded was needed. By 2002, the amount of oil spilled per barrel shipped had fallen 30% from the level three decades earlier.

Sources: Mark A. Cohen, "The Costs and Benefits of Oil Spill Prevention and Enforcement," Journal of Environmental Economics and Management 13(2) (June 1986): 167–188; Rick S. Kurtz, "Coastal Oil Pollution: Spills, Crisis, and Policy Change," Review of Policy Research, 21(2) (March 2004): 201–219; David S. Savage, "Justices Slash Exxon Valdez Verdict," Los Angeles Times, June 26, 2008, p. A1; and Edwin Unsworth, "Europe Gets Tougher on Aging Oil Tankers," Business Insurance, 36(48) (December 2, 2002): 33–34.

ANSWER TO TRY IT! PROBLEM

Here are the completed data table and the table showing total and marginal benefit and cost.



(b)

Hours studying economics	0		1		2		3		4		5
Total benefit	0		14		24		30		32		32
Marginal benefit		14		10		6		2		0	
Total cost	0		2		8		18		32		50
Marginal cost		2		6		10		14		18	

Ms. Phan maximizes her net benefit by reducing her time studying economics to 2 hours. The change in her expectations reduced the benefit and increased the cost of studying economics. The completed graph of marginal benefit and marginal cost is at the far left. Notice that answering the question using the marginal decision rule gives the same answer.

2. MAXIMIZING IN THE MARKETPLACE

LEARNING OBJECTIVES

1. **Explain what is meant by an efficient allocation of resources in an economy and describe the market conditions that must exist to achieve this goal.**
2. **Define consumer and producer surplus.**
3. **Discuss the relationship between efficiency and equity.**

In perhaps the most influential book in economics ever written, *An Inquiry into the Nature and Causes of the Wealth of Nations*, published in 1776, Adam Smith argued that the pursuit of self-interest in a marketplace would promote the general interest. He said resources would be guided, as if by an "invisible hand," to their best uses. That invisible hand was the marketplace.

Smith's idea was radical for its time; he saw that the seemingly haphazard workings of the marketplace could promote the common good. In this section, we will use the tools we have developed thus far to see the power of Smith's invisible hand. Efforts by individuals to maximize their own net benefit can maximize net benefit for the economy as a whole.

efficient

The allocation of resources when the net benefits of all economic activities are maximized.

When the net benefits of all economic activities are maximized, economists say the allocation of resources is **efficient**. This concept of efficiency is broader than the notion of efficient production that we encountered when discussing the production possibilities curve. There, we saw that the economy's factors of production would be efficient *in production* if they were allocated according to the principle of comparative advantage. That meant producing as much as possible with the factors of production available. The concept of an efficient allocation of resources incorporates production, as in that discussion, but it includes efficiency in the consumption of goods and services as well.

2.1 Achieving Efficiency

Imagine yourself arriving at the store to purchase some food. In your choice, you will weigh your own benefits and costs to maximize your net benefit. The farmers, the distributors, and the grocer have sought to maximize their net benefits as well. How can we expect that all those efforts will maximize net benefits for the economy as a whole? How can we expect the marketplace to achieve an efficient allocation of food, or of anything else?

One condition that must be met if the market's allocation is to be efficient is that the marketplace must be competitive or function as if it were. We will have a great deal more to say about competitive markets versus less competitive ones in subsequent chapters. For now, we can simply note that a competitive market is one with many buyers and sellers in each market and in which entry and exit are fairly easy. No one controls the price; the forces of demand and supply determine price.

The second condition that must hold if the market is to achieve an efficient allocation concerns property rights. We turn to that topic in the next section.

The Role of Property Rights

property rights

A set of rules that specify the ways in which an owner can use a resource.

A smoothly functioning market requires that producers possess property rights to the goods and services they produce and that consumers possess property rights to the goods and services they buy. **Property rights** are a set of rules that specify the ways in which an owner can use a resource.

Consider the tomato market. Farmers who grow tomatoes have clearly defined rights to their land and to the tomatoes they produce and sell. Distributors who purchase tomatoes from farmers and sell them to grocers have clear rights to the tomatoes until they sell them to grocers. The grocers who purchase the tomatoes retain rights to them until they sell them to consumers. When you buy a tomato, you have the exclusive right to its use.

A system of property rights forms the basis for all market exchange. Before exchange can begin, there must be a clear specification of who owns what. The system of property rights must also show what purchasers are acquiring when they buy rights to particular resources. Because property rights must exist if exchange is to occur, and because exchange is the process through which economic efficiency is achieved, a system of property rights is essential to the efficient allocation of resources.

Imagine what would happen in the market for tomatoes if property rights were not clearly defined. Suppose, for example, that grocers could not legally prevent someone from simply grabbing some tomatoes and leaving without paying for them. If that were the case, grocers would not be likely to offer tomatoes for sale. If it were the case for all grocery items, there would not be grocery stores at all.

Although property rights vary for different resources, two characteristics are required if the marketplace is to achieve an efficient allocation of resources:

exclusive property right

A property right that allows its owner to prevent others from using the resource.

transferable property right

A property right that allows the owner of a resource to sell or lease it to someone else.

1. Property rights must be exclusive. An **exclusive property right** is one that allows its owner to prevent others from using the resource. The owner of a house, for example, has the right to exclude others from the use of the house. If this right did not exist, ownership would have little value; it is not likely that the property could be exchanged in a market. And the inability to sell property would limit the incentive of owners to maintain it.
2. Property rights must be transferable. A **transferable property right** is one that allows the owner of a resource to sell or lease it to someone else. In the absence of transferability, no exchange could occur.

Markets and the Efficiency Condition

A competitive market with well-defined and transferable property rights satisfies the **efficiency condition**. If met, we can assume that the market's allocation of resources will be efficient.

efficient

The allocation of resources when the net benefits of all economic activities are maximized.

Consider again your purchase of tomatoes. Suppose the curves of demand and supply for tomatoes are those given in Figure 6.10; the equilibrium price equals $1.50 per pound. Suppose further that the market satisfies the efficiency condition. With that assumption, we can relate the model of demand and supply to our analysis of marginal benefits and costs.

The demand curve tells us that the last pound of tomatoes was worth $1.50; we can think of that as the marginal benefit of the last pound of tomatoes since that is how much consumers were willing to pay. We can say that about any price on a market demand curve; a demand curve can be considered as a marginal benefit curve. Similarly, the supply curve can be considered the marginal cost curve. In the case of the tomato market, for example, the price tells us that the marginal cost of producing the last pound of tomatoes is $1.50. This marginal cost is considered in the economic sense—other goods and services worth $1.50 were not produced in order to make an additional pound of tomatoes available.

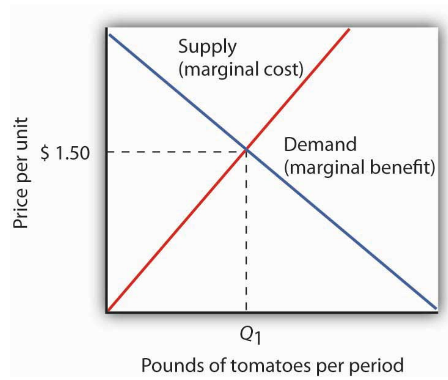
On what basis can we presume that the price of a pound of tomatoes equals its marginal cost? The answer lies in our marginal decision rule. Profit-maximizing tomato producers will produce more tomatoes as long as their marginal benefit exceeds their marginal cost. What is the marginal benefit to a producer of an extra pound of tomatoes? It is the price that the producer will receive. What is the marginal cost? It is the value that must be given up to produce an extra pound of tomatoes.

Producers maximize profit by expanding their production up to the point at which their marginal cost equals their marginal benefit, which is the market price. The price of $1.50 thus reflects the marginal cost to society of making an additional pound of tomatoes available.

At the equilibrium price and output of tomatoes, then, the marginal benefit of tomatoes to consumers, as reflected by the price they are willing to pay, equals the marginal cost of producing tomatoes. Where marginal benefit equals marginal cost, net benefit is maximized. The equilibrium quantity of tomatoes, as determined by demand and supply, is efficient.

FIGURE 6.10 Demand and Supply and the Efficiency Condition

In a competitive market with exclusive and transferable property rights, such as the market for tomatoes, the efficiency condition is met. Buyers and sellers are faced with all of the relevant benefits and costs, and the equilibrium price equals the marginal cost to society of producing that good, here $2.50 per pound. We can interpret the market demand and supply curve as marginal benefit and marginal cost curves, respectively.



2.2 Producer and Consumer Surplus

Think about the last thing you purchased. You bought it because you expected that its benefits would exceed its opportunity cost; you expected that the purchase would make you better off. The seller sold it to you because he or she expected that the money you paid would be worth more than the value of keeping the item. The seller expected to be better off as a result of the sale. Exchanges in the marketplace have a remarkable property: Both buyers and sellers expect to emerge from the transaction better off.

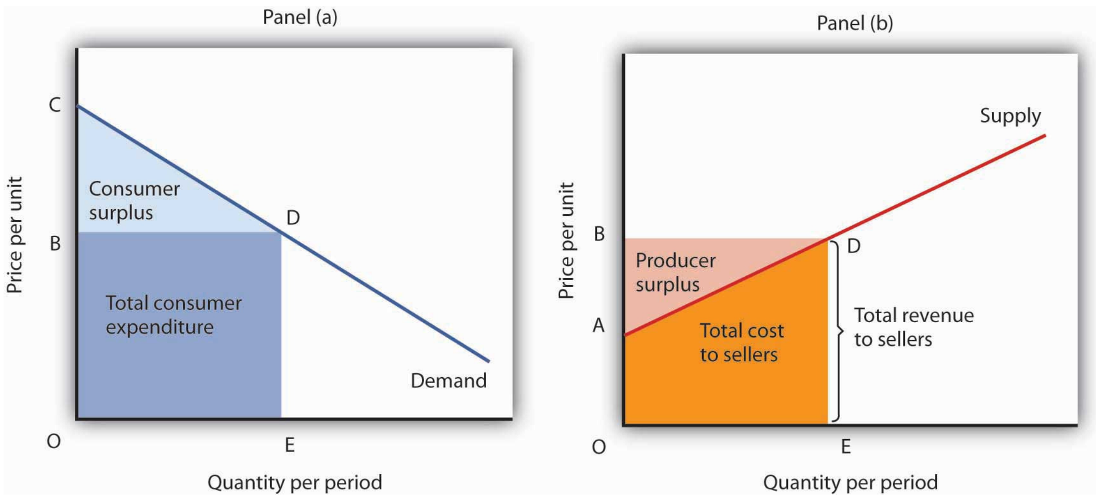
Panel (a) of Figure 6.11 shows a market demand curve for a particular good. Suppose the price equals OB and the quantity equals OE. The area under the demand curve over the range of quantities from the origin at O to the quantity at E equals the total benefit of consuming OE units of the good. It is the area OCDE. Consumers pay for this benefit; their total expenditures equal the rectangle OBDE, which is the dark shaded region in the graph. Because the total benefits exceed total expenditures, there is a consumer surplus given by the triangle BCD. **Consumer surplus** is the amount by which the total benefits to consumers from consuming a good exceed their total expenditures on the good.

consumer surplus

The amount by which the total benefits to consumers from consuming a good exceed their total expenditures on the good.

FIGURE 6.11 Consumer and Producer Surplus

Consumer surplus [Panel (a)] measures the difference between total benefit of consuming a given quantity of output and the total expenditures consumers pay to obtain that quantity. Here, total benefits are given by the shaded area OCDE; total expenditures are given by the rectangle OBDE. The difference, shown by the triangle BCD, is consumer surplus. Producer surplus [Panel b)] measures the difference between total revenue received by firms at a given quantity of output and the total cost of producing it. Here, total revenue is given by the rectangle OBDE, and total costs are given by the area OADE. The difference, shown by the triangle ABD is producer surplus.



Now consider the sellers' side of transactions. Panel (b) of Figure 6.11 shows a market supply curve; recall that it gives us marginal cost. Suppose the market price equals OB and quantity supplied is OE; those are the same values we had in Panel (a). The price times the quantity equals the total revenue received by sellers. It is shown as the shaded rectangle OBDE. The total revenue received by sellers equals total expenditures by consumers.

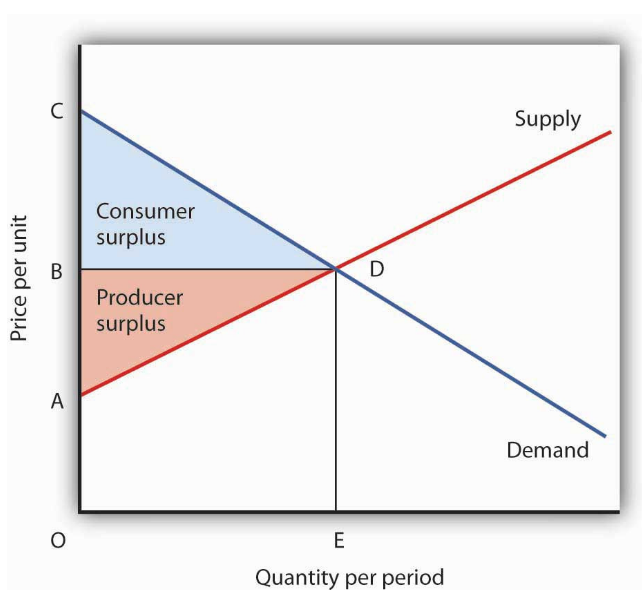
producer surplus

The difference between the total revenue received by sellers and their total cost.

The total cost to sellers is the area under the marginal cost curve; it is the area OADE. That cost is less than revenue. The difference between the total revenue received by sellers and their total cost is called **producer surplus**. In Panel (b) it is the light-shaded triangle ABD.

FIGURE 6.12 Net Benefit: The Sum of Consumer and Producer Surplus

The sum of consumer surplus and producer surplus measures the net benefit to society of any level of economic activity. Net benefit is maximized when production and consumption are carried out at the level where the demand and supply curves intersect. Here, the net benefit to society equals the area ACD. It is the sum of consumer surplus, BCD, and producer surplus, ABD.



We put the demand and supply curves of Figure 6.11 Panels (a) and (b) together in Figure 6.12. The intersection of the two curves determines the equilibrium price, OB, and the equilibrium quantity, OE. The shaded regions give us consumer and producer surplus. The sum of these two surpluses is net benefit. This net benefit is maximized where the demand and supply curves intersect.

2.3 Efficiency and Equity

Consumer demands are affected by incomes. Demand, after all, reflects ability as well as willingness to pay for goods and services. The market will be more responsive to the preferences of people with high incomes than to those of people with low incomes.

In a market that satisfies the efficiency condition, an efficient allocation of resources will emerge from any particular distribution of income. Different income distributions will result in different, but still efficient, outcomes. For example, if 1% of the population controls virtually all the income, then the market will efficiently allocate virtually all its production to those same people.

What is a fair, or equitable, distribution of income? What is an unfair distribution? Should everyone have the same income? Is the current distribution fair? Should the rich have less and the poor have more? Should the middle class have more? Equity is very much in the mind of the observer. What may seem equitable to one person may seem inequitable to another. There is, however, no test we can apply to determine whether the distribution of income is or is not equitable. That question requires a normative judgment.

Determining whether the allocation of resources is or is not efficient is one problem. Determining whether the distribution of income is fair is another. The governments of all nations act in some way to redistribute income. That fact suggests that people generally have concluded that leaving the distribution of income solely to the market would not be fair and that some redistribution is desirable. This may take the form of higher taxes for people with higher incomes than for those with lower incomes. It may take the form of special programs, such as welfare programs, for low-income people.

Whatever distribution society chooses, an efficient allocation of resources is still preferred to an inefficient one. Because an efficient allocation maximizes net benefits, the gain in net benefits could be distributed in a way that leaves all people better off than they would be at any inefficient allocation. If an efficient allocation of resources seems unfair, it must be because the distribution of income is unfair.

KEY TAKEAWAYS

- In a competitive system in which the interaction of demand and supply determine prices, the corresponding demand and supply curves can be considered marginal benefit and marginal cost curves, respectively.
- An efficient allocation of resources is one that maximizes the net benefit of each activity. We expect it to be achieved in markets that satisfy the efficiency condition, which requires a competitive market and well-defined, transferable property rights.
- Consumer surplus is the amount by which the total benefit to consumers from some activity exceeds their total expenditures for it.
- Producer surplus is the amount by which the total revenues of producers exceed their total costs.
- An inequitable allocation of resources implies that the distribution of income and wealth is inequitable. Judgments about equity are normative judgments.

TRY IT!

Draw hypothetical demand and supply curves for a typical product, say coffee. Now show the areas of consumer and producer surplus. Under what circumstances is the market likely to be efficient?

Case in Point: Saving the Elephant Through Property Rights

The African elephant, the world's largest land mammal, seemed to be in danger of extinction in the 20th century. The population of African elephants fell from 1.3 million in 1979 to 543,000 in 1994. The most dramatic loss of elephants came in Kenya, where the population fell from 167,000 early in the 1970s to about 26,000 in 1997, according to the World Wildlife Fund. To combat the slaughter, an international agreement, the Convention on International Trade in Endangered Species of Wild Flora and Fauna (CITES), went into effect in 1989. It banned the sale of ivory.

Despite CITES and armed patrols with orders to shoot poachers on sight, the poachers continued to operate in Kenya, killing roughly 200 elephants per day. The elephants were killed for their ivory; the tusks from a single animal could be sold for $2,000 in the black market—nearly double the annual per capita income in Kenya.

Several African nations, however, have taken a radically different approach. They have established exclusive, transferable property rights in licenses to hunt elephants. In each of these nations, elephant populations have increased. These nations include Botswana, Namibia, South Africa, Tanzania, Zambia, and Zimbabwe. In Botswana, for example, the elephant population increased from 20,000 in 1981 to 80,000 in 2000. Zimbabwe increased its elephant population from 30,000 in 1978 to nearly 90,000 in 2000.

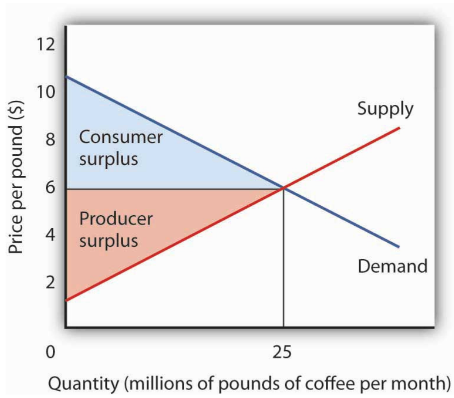
Professors Michael A. McPherson and Michael L. Nieswiadomy of the University of North Texas have done a statistical analysis of the determinants of elephant populations in 35 African nations. They found that elephant populations increased in nations that had (a) established exclusive, transferable property rights in licenses to hunt elephants and (b) had stable political systems. Conversely, elephant populations declined in countries that had failed to establish property rights and that had unstable political systems.

The same appears to be true of the white rhinoceros, a creature whose horns are highly valued in Asia as an aphrodisiac. South Africa sells permits to hunt the creatures for $25,000 per animal. Its rhinoceros herd has increased from 20 in 1900 to more than 7,000 by the late 1990s.

There is no "secret" to the preservation of species. Establishing clearly defined, transferable property rights virtually assures the preservation of species. Whether it be buffaloes, rhinoceroses, or elephants, property rights establish a market, and that market preserves species.

Sources: Lisa Grainger, "Are They Safe in Our Hands?" The Times of London (July 16, 1994): p. 18; Michael A. McPherson and Michael L. Nieswiadomy, "African Elephants: The Effect of Property Rights and Political Stability," Contemporary Economic Policy, 18(1) (January 2000): 14–26; "Tusks and Horns and Conservationists," The Economist, 343(8019) (May 31, 1997): 44.

ANSWER TO TRY IT! PROBLEM



On the assumption that the coffee market is competitive and that it is characterized by well-defined exclusive and transferable property rights, the coffee market meets the efficiency condition. That means that the allocation of resources shown at the equilibrium will be the one that maximizes the net benefit of all activities. The net benefit is shared by coffee consumers (as measured by consumer surplus) and coffee producers (as measured by producer surplus).

3. MARKET FAILURE

LEARNING OBJECTIVES

1. **Explain what is meant by market failure and the conditions that may lead to it.**
2. **Distinguish between private goods and public goods and relate them to the free rider problem and the role of government.**
3. **Explain the concepts of external costs and benefits and the role of government intervention when they are present.**
4. **Explain why a common property resource is unlikely to be allocated efficiently in the marketplace.**

Private decisions in the marketplace may not be consistent with the maximization of the net benefit of a particular activity. The failure of private decisions in the marketplace to achieve an efficient allocation of scarce resources is called **market failure**. Markets will not generate an efficient allocation of resources if they are not competitive or if property rights are not well defined and fully transferable. Either condition will mean that decision makers are not faced with the marginal benefits and costs of their choices.

market failure

The failure of private decisions in the marketplace to achieve an efficient allocation of scarce resources.

Think about the drive that we had you take at the beginning of this chapter. You faced some, but not all, of the opportunity costs involved in that choice. In particular, your choice to go for a drive would increase air pollution and might increase traffic congestion. That means that, in weighing the marginal benefits and marginal costs of going for a drive, not all of the costs would be counted. As a result, the net benefit of the allocation of resources such as the air might not be maximized.

3.1 Noncompetitive Markets

The model of demand and supply assumes that markets are competitive. No one in these markets has any power over the equilibrium price; each consumer and producer takes the market price as given and responds to it. Under such conditions, price is determined by the intersection of demand and supply.

In some markets, however, individual buyers or sellers are powerful enough to influence the market price. In subsequent chapters, we will study cases in which producers or consumers are in a position to affect the prices they charge or must pay, respectively. We shall find that when individual firms or groups of firms have market power, which is the ability to change the market price, the price will be distorted—it will not equal marginal cost.

3.2 Public Goods

public good

A good for which the cost of exclusion is prohibitive and for which the marginal cost of an additional user is zero.

private good

A good for which exclusion is possible and for which the marginal cost of another user is positive.

Some goods are unlikely to be produced and exchanged in a market because of special characteristics of the goods themselves. The benefits of these goods are such that exclusion is not feasible. Once they are produced, anyone can enjoy them; there is no practical way to exclude people who have not paid for them from consuming them. Furthermore, the marginal cost of adding one more consumer is zero. A good for which the cost of exclusion is prohibitive and for which the marginal cost of an additional user is zero is a **public good**. A good for which exclusion is possible and for which the marginal cost of another user is positive is a **private good**.

National defense is a public good. Once defense is provided, it is not possible to exclude people who have not paid for it from its consumption. Further, the cost of an additional user is zero—an army does not cost any more if there is one more person to be protected. Other examples of public goods include law enforcement, fire protection, and efforts to preserve species threatened with extinction.

Free Riders

Suppose a private firm, Terror Alert, Inc., develops a completely reliable system to identify and intercept 98% of any would-be terrorists that might attempt to enter the United States from anywhere in the world. This service is a public good. Once it is provided, no one can be excluded from the system's protection on grounds that he or she has not paid for it, and the cost of adding one more person to the group protected is zero. Suppose that the system, by eliminating a potential threat to U.S. security, makes the average person in the United States better off; the benefit to each household from the added security is worth $40 per month (about the same as an earthquake insurance premium). There are roughly 113 million households in the United States, so the total benefit of the system is $4.5 billion per month. Assume that it will cost Terror Alert, Inc., $1 billion per month to operate. The benefits of the system far outweigh the cost.

Suppose that Terror Alert installs its system and sends a bill to each household for $20 for the first month of service—an amount equal to half of each household's benefit. If each household pays its bill, Terror Alert will enjoy a tidy profit; it will receive revenues of more than $2.25 billion per month.

free riders

People or firms that consume a public good without paying for it.

But will each household pay? Once the system is in place, each household would recognize that it will benefit from the security provided by Terror Alert whether it pays its bill or not. Although some households will voluntarily pay their bills, it seems unlikely that very many will. Recognizing the opportunity to consume the good without paying for it, most would be free riders. **Free riders** are people or firms that consume a public good without paying for it. Even though the total benefit of the system is $4.5 billion, Terror Alert will not be faced by the marketplace with a signal that suggests that the system is worthwhile. It is unlikely that it will recover its cost of $1 billion per month. Terror Alert is not likely to get off the ground.

The bill for $20 from Terror Alert sends the wrong signal, too. An efficient market requires a price equal to marginal cost. But the marginal cost of protecting one more household is zero; adding one more household adds nothing to the cost of the system. A household that decides not to pay Terror Alert anything for its service is paying a price equal to its marginal cost. But doing that, being a free rider, is precisely what prevents Terror Alert from operating.

Because no household can be excluded and because the cost of an extra household is zero, the efficiency condition will not be met in a private market. What is true of Terror Alert, Inc., is true of public goods in general: they simply do not lend themselves to private market provision.

Public Goods and the Government

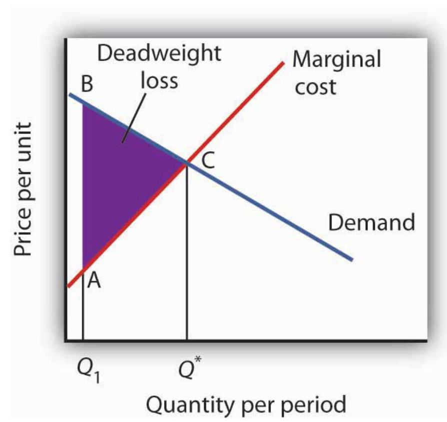
Because many individuals who benefit from public goods will not pay for them, private firms will produce a smaller quantity of public goods than is efficient, if they produce them at all. In such cases, it may be desirable for government agencies to step in. Government can supply a greater quantity of the good by direct provision, by purchasing the public good from a private agency, or by subsidizing consumption. In any case, the cost is financed through taxation and thus avoids the free-rider problem.

Most public goods are provided directly by government agencies. Governments produce national defense and law enforcement, for example. Private firms under contract with government agencies produce some public goods. Park maintenance and fire services are public goods that are sometimes produced by private firms. In other cases, the government promotes the private consumption or production of public goods by subsidizing them. Private charitable contributions often support activities that are public goods; federal and state governments subsidize these by allowing taxpayers to reduce their tax payments by a fraction of the amount they contribute.

While the market will produce some level of public goods in the absence of government intervention, we do not expect that it will produce the quantity that maximizes net benefit. Figure 6.15 illustrates the problem. Suppose that provision of a public good such as national defense is left entirely to private firms. It is likely that some defense services would be produced; suppose that equals Q_1 units per period. This level of national defense might be achieved through individual contributions. But it is very unlikely that contributions would achieve the correct level of defense services. The efficient quantity occurs where the demand, or marginal benefit, curve intersects the marginal cost curve, at Q^*. The deadweight loss is the shaded area ABC; we can think of this as the net benefit of government intervention to increase the production of national defense from Q_1 up to the efficient quantity, Q^*.

FIGURE 6.15 Public Goods and Market Failure

Because free riders will prevent firms from being able to require consumers to pay for the benefits received from consuming a public good, output will be less than the efficient level. In the case shown here, private donations achieved a level of the public good of Q_1 per period. The efficient level is Q^*. The deadweight loss is shown by the triangle ABC.



Heads Up!

Note that the definitions of public and private goods are based on characteristics of the goods themselves, not on whether they are provided by the public or the private sector. Postal services are a private good provided by the public sector. The fact that these goods are produced by a government agency does not make them a public good.

3.3 External Costs and Benefits

Suppose that in the course of production, the firms in a particular industry generate air pollution. These firms thus impose costs on others, but they do so outside the context of any market exchange—no agreement has been made between the firms and the people affected by the pollution. The firms thus will not be faced with the costs of their action. A cost imposed on others outside of any market exchange is an **external cost**.

external cost

A cost imposed on others outside of any market exchange.

We saw an example of an external cost in our imaginary decision to go for a drive. Here is another: violence on television, in the movies, and in video games. Many critics argue that the violence that pervades these media fosters greater violence in the real world. By the time a child who spends the average amount of time watching television finishes elementary school, he or she will have seen 100,000 acts of violence, including 8,000 murders, according to the American Psychological Association. Thousands of studies of the relationship between violence in the media and behavior have concluded that there is a link between watching violence and violent behaviors. Video games are a major element of the problem, as young children now spend hours each week playing them. Fifty percent of fourth-grade graders say that their favorite video games are the “first person shooter” type.[1]

Any tendency of increased violence resulting from increased violence in the media constitutes an external cost of such media. The American Academy of Pediatrics reported in 2001 that homicides were the fourth leading cause of death among children between the ages of 10 and 14 and the second leading cause of death for people aged 15 to 24 and has recommended a reduction in exposure to media violence.[2] It seems reasonable to assume that at least some of these acts of violence can be considered an external cost of violence in the media.

An action taken by a person or firm can also create benefits for others, again in the absence of any market agreement; such a benefit is called an **external benefit**. A firm that builds a beautiful building generates benefits to everyone who admires it; such benefits are external.

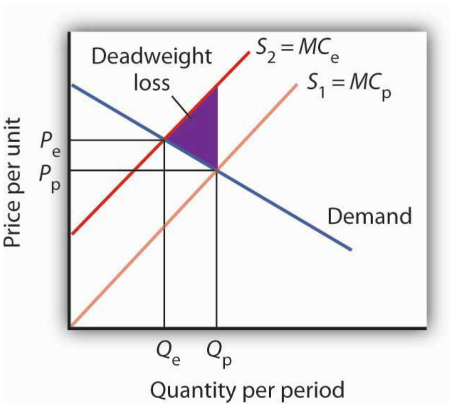
external benefit

An action taken by a person or firm that creates benefits for others in the absence of any market agreement.

External Costs and Efficiency

FIGURE 6.16 External Costs

When firms in an industry generate external costs, the supply curve S_1 reflects only their private marginal costs, MC_p. Forcing firms to pay the external costs they impose shifts the supply curve to S_2, which reflects the full marginal cost of the firms' production, MC_e. Output is reduced and price goes up. The deadweight loss that occurs when firms are not faced with the full costs of their decisions is shown by the shaded area in the graph.



The case of the polluting firms is illustrated in Figure 6.16. The industry supply curve S_1 reflects private marginal costs, MC_p. The market price is P_p for a quantity Q_p. This is the solution that would occur if firms generating external costs were not forced to pay those costs. If the external costs generated by the pollution were added, the new supply curve S_2 would reflect higher marginal costs, MC_e. Faced with those costs, the market would generate a lower equilibrium quantity, Q_e. That quantity would command a higher price, P_e. The failure to confront producers with the cost of their pollution means that consumers do not pay the full cost of the good they are purchasing. The level of output and the level of pollution are therefore higher than would be economically efficient. If a way could be found to confront producers with the full cost of their choices, then consumers would be faced with a higher cost as well. Figure 6.16 shows that consumption would be reduced to the efficient level, Q_e, at which demand and the full marginal cost curve (MC_e) intersect. The deadweight loss generated by allowing the external cost to be generated with an output of Q_p is given as the shaded region in the graph.

External Costs and Government Intervention

If an activity generates external costs, the decision makers generating the activity will not be faced with its full costs. Agents who impose these costs will carry out their activities beyond the efficient level; those who consume them, facing too low a price, will consume too much. As a result, producers and consumers will carry out an excessive quantity of the activity. In such cases, government may try to intervene to reduce the level of the activity toward the efficient quantity. In the case shown in Figure 6.16, for example, firms generating an external cost have a supply curve S_1 that reflects their private marginal costs, MC_p. A per-unit pollution fee imposed on the firms would increase their marginal costs to MC_e, thus shifting the supply curve to S_2, and the efficient level of production would emerge. Taxes or other restrictions may be imposed on the activity that generates the external cost in an effort to confront decision makers with the costs that they are imposing. In many areas, firms and consumers that pollute rivers and lakes are required to pay fees based on the amount they pollute. Firms in many areas are required to purchase permits in order to pollute the air; the requirement that permits be purchased serves to confront the firms with the costs of their choices.

Another approach to dealing with problems of external costs is direct regulation. For example, a firm may be ordered to reduce its pollution. A person who turns his or her front yard into a garbage dump may be ordered to clean it up. Participants at a raucous party may be told to be quiet. Alternative ways of dealing with external costs are discussed later in the text.

3.4 Common Property Resources

common property resource

Resources for which no property rights have been defined.

Common property resources[3] are resources for which no property rights have been defined. The difficulty with common property resources is that individuals may not have adequate incentives to engage in efforts to preserve or protect them. Consider, for example, the relative fates of cattle and buffalo in the United States in the nineteenth century. Cattle populations increased throughout the century, while the buffalo nearly became extinct. The chief difference between the two animals was that exclusive property rights existed for cattle but not for buffalo.

Owners of cattle had an incentive to maintain herd sizes. A cattle owner who slaughtered all of his or her cattle without providing for replacement of the herd would not have a source of future income. Cattle owners not only maintained their herds but also engaged in extensive efforts to breed high-quality livestock. They invested time and effort in the efficient management of the resource on which their livelihoods depended.

Buffalo hunters surely had similar concerns about the maintenance of buffalo herds, but they had no individual stake in doing anything about them—the animals were a common property resource. Thousands of individuals hunted buffalo for a living. Anyone who cut back on hunting in order to help to preserve the herd would lose income—and face the likelihood that other hunters would go on hunting at the same rate as before.

Today, exclusive rights to buffalo have been widely established. The demand for buffalo meat, which is lower in fat than beef, has been increasing, but the number of buffalo in the United States is rising rapidly. If buffalo were still a common property resource, that increased demand, in the absence of other restrictions on hunting of the animals, would surely result in the elimination of the animal. Because there are exclusive, transferable property rights in buffalo and because a competitive market brings buyers and sellers of buffalo and buffalo products together, we can be reasonably confident in the efficient management of the animal.

When a species is threatened with extinction, it is likely that no one has exclusive property rights to it. Whales, condors, grizzly bears, elephants in Central Africa—whatever the animal that is threatened—are common property resources. In such cases a government agency may impose limits on the killing of the animal or destruction of its habitat. Such limits can prevent the excessive private use of a common property resource. Alternatively, as was done in the case of the buffalo, private rights can be established, giving resource owners the task of preservation.

KEY TAKEAWAYS

- Public sector intervention to increase the level of provision of public goods may improve the efficiency of resource allocation by overcoming the problem of free riders.
- Activities that generate external costs are likely to be carried out at levels that exceed those that would be efficient; the public sector may seek to intervene to confront decision makers with the full costs of their choices.
- Some private activities generate external benefits.
- A common property resource is unlikely to be allocated efficiently in the marketplace.

TRY IT!

The manufacture of memory chips for computers generates pollutants that generally enter rivers and streams. Use the model of demand and supply to show the equilibrium price and output of chips. Assuming chip manufacturers do not have to pay the costs these pollutants impose, what can you say about the efficiency of the quantity of chips produced? Show the area of deadweight loss imposed by this external cost. Show how a requirement that firms pay these costs as they produce the chips would affect the equilibrium price and output of chips. Would such a requirement help to satisfy the efficiency condition? Explain.

Case in Point: Externalities and Smoking

Smokers impose tremendous costs on themselves. Based solely on the degree to which smoking shortens their life expectancy, which is by about six years, the cost per pack is $35.64. That cost, of course, is a private cost. In addition to that private cost, smokers impose costs on others. Those external costs come in three ways. First, they increase health-care costs and thus increase health insurance premiums. Second, smoking causes fires that destroy more than $300 million worth of property each year. Third, more than 2,000 people die each year as a result of "secondhand" smoke. A 1989 study by the RAND Corporation estimated these costs at $0.53 per pack.

In an important way, however, smokers also generate external benefits. They contribute to retirement programs and to Social Security, then die sooner than nonsmokers. They thus subsidize the retirement programs of the rest of the population. According to the RAND study, that produces an external benefit of $0.24 per pack, leaving a net external cost of $0.29 per pack. Given that state and federal excise taxes averaged $0.37 in 1989, the RAND researchers concluded that smokers more than paid their own way.

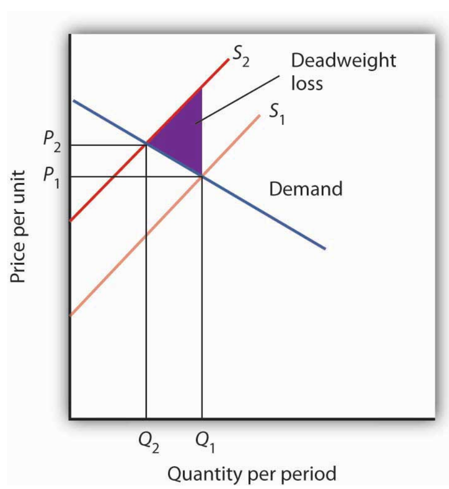
Economists Jonathan Gruber of the Massachusetts Institute of Technology and Botond Koszegi of the University of California at Berkeley have suggested that, in the case of people who consume "addictive bads" such as cigarettes, an excise tax on cigarettes of as much as $4.76 per pack may improve the welfare of smokers.

They base their argument on the concept of "time inconsistency," which is the theory that smokers seek the immediate gratification of a cigarette and then regret their decision later. Professors Gruber and Koszegi argue that higher taxes would serve to reduce the quantity of cigarettes demanded and thus reduce behavior that smokers would otherwise regret. Their argument is that smokers impose "internalities" on themselves and that higher taxes would reduce this.

Where does this lead us? If smokers are "rationally addicted" to smoking, i.e., they have weighed the benefits and costs of smoking and have chosen to smoke, then the only problem for public policy is to assure that smokers are confronted with the external costs they impose. In that case, the problem is solved: through excise taxes, smokers more than pay their own way. But, if the decision to smoke is an irrational one, it may be improved through higher excise taxes on smoking.

Sources: Jonathan Gruber and Botond Koszegi, "A Theory of Government Regulation of Addictive Bads: Optimal Tax Levels and Tax Incidence for Cigarette Excise Taxation," NBER Working Paper 8777, February 2002; Willard G. Manning et al., "The Taxes of Sin: Do Smokers and Drinkers Pay Their Way?" Journal of the American Medical Association, 261 (March 17, 1989): 1604–1609.

ANSWER TO TRY IT! PROBLEM



In the absence of any regulation, chip producers are not faced with the costs of the pollution their operations generate. The market price is thus P_1 and the quantity Q_1. The efficiency condition is not met; the price is lower and the quantity greater than would be efficient. If producers were forced to face the cost of their pollution as well as other production costs, the supply curve would shift to S_2, the price would rise to P_2, and the quantity would fall to Q_2. The new solution satisfies the efficiency condition.

4. REVIEW AND PRACTICE

Summary

Economists insist that individuals do not make choices willy-nilly. Rather, economists assume that individuals make choices in a purposeful way, one that seeks the maximum value for some objective. We assume that consumers seek to maximize utility and that firms seek to maximize profits.

Whatever is being maximized, choices are based on the marginal decision rule. Following this rule results in an allocation that achieves the greatest degree of utility or profit possible.

If utility- and profit-maximizing choices are made in the context of a price system that confronts decision makers with all of the costs and all of the benefits of their choices, the allocation of resources will be efficient. An efficient allocation is one that maximizes the net benefit of every activity. The concepts of consumer and producer surplus show us how this net benefit is shared. Equity is a separate issue, one that calls for a normative evaluation of the fairness of the distribution of income.

The allocation of resources will be inefficient in the absence of competitive markets. It will also be inefficient if property rights are not exclusive and transferable. These two conditions break down when there are public goods, common property resources, or external benefits or costs. In each of these cases, public sector intervention may improve the efficiency of resource allocation. When a market fails to achieve the efficient solution, net benefit falls short of the maximum possible. Deadweight loss is the amount by which net benefit falls below the net benefit possible at the efficient solution.

CONCEPT PROBLEMS

1. What is achieved by selecting the quantity of an activity at which marginal benefit equals marginal cost?
2. Suppose the marginal benefit of an activity exceeds the marginal cost. What does the marginal decision rule say a maximizing decision maker will do?
3. Suppose you are a discus hurler and your goal is to maximize the distance you achieve. You "produce" discus hurls by practicing. The total benefit of practice is distance achieved, and the input that achieves this distance is hours of practice. Describe the total benefit curve of practice. What point on the curve would you select?
4. This chapter argues that consumers maximize utility and firms maximize profits. What do you suppose each of the following might be presumed to maximize?
    a. A minister or rabbi
    b. A United States Senator
    c. The manager of a major league baseball team
    d. The owner of a major league baseball team
    e. The director of a charitable organization
5. For each of the following goods, indicate whether exclusive, transferable property rights exist and whether the good poses a problem for public policy. If it does, does the problem relate to a problem of property rights?
    a. Clean air
    b. Tomatoes
    c. Housing
    d. Blue whales
6. The dry-cleaning industry is a major source of air pollution. What can you conclude about the price and output of dry-cleaning services?
7. Economists often recommend that polluters such as dry-cleaning establishments be charged fees for the pollution they emit. Critics of this idea respond that the establishments would simply respond by passing these charges on to their customers, leaving the level of pollution unchanged. Comment on this objection.
8. Government agencies often require that children be inoculated against communicable diseases such as polio and measles. From the standpoint of economic efficiency, is there any justification for such a requirement?
9. Which of the following goods or services are public? Why or why not?
    a. Libraries
    b. Fire protection
    c. Television programs
    d. Health care
    e. Water for household consumption
10. If a village in Botswana is granted several licenses to kill elephants, how does this give it an incentive to preserve elephants and increase the size of the herd? How does the international ban on ivory sales affect the incentive in Botswana to preserve the elephant?
11. The number of fish caught in the ocean has fallen in recent years partly as a result of more intensive fishing efforts and the use of more sophisticated equipment. Fish in the ocean are a common property resource. How might this fact be related to declining fish catches? How do you think this drop in the catch affects the price of seafood?

NUMERICAL PROBLEMS

1. Joe Higgins is thinking about how much time to spend studying for a biology exam tomorrow. Using "utility units" he measures the benefits and costs of study; his calculations are shown in the following table.

Hours spent studying	0	1	2	3	4	5	6
Total benefit	0	100	180	240	280	300	300
Total cost	0	50	100	150	200	250	300
Net benefit							

   a. Fill in the fourth row for net benefit in the table. Use the midpoint convention to emphasize that the net benefit is a marginal value showing the gain as hours spent increase by one-hour increments.
   b. Using a graph similar to Panel (a) of Figure 6.1 show the marginal benefit curve and verify that the area under the curve at 3 hours of study corresponds to the total benefit of that much study. (Hint: Remember that marginal values are plotted at the midpoints of the corresponding intervals on the horizontal axis.)
   c. Use a graph similar to Panel (b) of Figure 6.1 to show the marginal cost curve and verify that the area under the curve at 3 hours of study corresponds to the total cost of that much study.
   d. Use a graph similar to Panel (a) of Figure 6.6 to combine the marginal benefit and marginal cost curves you drew in parts (a) and (b).
   e. Based on the marginal decision rule, how many hours should Joe spend studying for his biology exam?

2. Now suppose some friends of Joe's call to say they are having a party tonight. Joe calculates that the party is now his best alternative to study, and he increases his estimate of the cost of each hour of study. One hour of study now costs 70; two hours cost 140; three hours 210, four hours 280; five hours 350; and six hours 470.
   a. Draw the new marginal benefit and marginal cost curves as in Problem 1, part (d):
   b. Based on the marginal decision rule, identify the new solution that maximizes the net benefit of study time.

3. The local gasoline market in a particular city has demand and supply curves given by the following data. (All quantities are in millions of gallons per month.)

Price per gallon	$1.00	$1.50	$2.00	$2.50	$3.00	$3.50	$4.00
Quantity demanded	6	5	4	3	2	1	0
Quantity supplied	0	1	2	3	4	5	6

   a. Plot the demand and supply curves, and determine the equilibrium price and quantity.
   b. Show the areas of consumer and producer surplus.
   c. Now suppose that the community determines that each gallon of gasoline consumed imposes $0.50 in pollution costs. Accordingly, a $0.50-per-gallon tax is imposed. The tax is imposed on sellers of gasoline, and it has the effect of increasing by $0.50 the price required to induce the quantities supplied in the table. For example, a price of $2.00 is now required for a quantity of 1 million gallons to be supplied each month. Plot the new supply curve.
   d. Approximate the new equilibrium price and output.
   e. Does the price increase by the full amount of the tax? If not, explain why.
   f. Would your answer be different if the demand for gasoline were perfectly inelastic?

4. The flu vaccination market has the demand and supply curves given by the following data. (All quantities are in thousands.)

Price per vaccination	$10	$15	$20	$25	$30
Quantity demanded	90	80	70	60	50
Quantity supplied	50	60	70	80	90

   a. Plot the demand and supply curves, and determine the equilibrium price and quantity.
   b. Show the areas of consumer and producer surplus.

c. Now suppose that each vaccination given generates an external benefit, as those who do not get vaccinated are less likely to get the flu when others do get vaccinated. As a result, suppliers receive a $10 subsidy from the government for each vaccine. For example, if consumers pay $10 per vaccination, suppliers receive $20, so only $10 from consumers is required to induce suppliers to offer 70,000 vaccinations per month. Plot the new supply curve.

d. Determine the new equilibrium price and quantity.

e. Does the price fall by the full amount of the subsidy? If not, explain why.

f. What is the total amount that consumers now pay for the new equilibrium quantity of vaccinations?

g. What is the total subsidy that suppliers receive from the government at the new equilibrium quantity of vaccinations?

5. Given the following information about the supply of and demand for apples:

Price per pound	Quantity demanded (pounds per month)	Quantity Supplied (pounds per month
$0.50	12,000	0
0.75	10,000	2,000
1.00	8,000	4,000
1.25	6,000	6,000
1.50	4,000	8,000
1.75	2,000	10,000
2.00	0	12,000

a. Draw a graph similar to Figure 6.12

b. Assuming the market for apples meets the efficiency condition, show the equilibrium price and quantity that maximizes net benefit to society.

c. Identify the area of consumer surplus and the area of producer surplus.

ENDNOTES

1. See Report of the Committee on Commerce, Science, and Transportation, *Children's Protection From Violent Programming Act*, Senate Report 106–509 (October 26, 2000), Washington, D.C.: U.S. Government Printing Office, 2000, and Michael Rich, "Violent Video Games Testimony," Chicago City Council, October 30, 2000, at http://www.aap.org/advocacy/rich-videogameviolence.pdf.
2. Mark Rosenberg, "Successful State Strategies," Adolescent Health Leadership Forum, December 6, 2003, at http://www.aap.org/advocacy/ahproject/AHLSuccessfulStateStrategiesMRosenberg.pps.
3. Common property resources are sometimes referred to as open access resources.

CHAPTER 8
Production and Cost

START UP: STREET CLEANING AROUND THE WORLD

It is dawn in Shanghai, China. Already thousands of Chinese are out cleaning the city's streets. They are using brooms.

On the other side of the world, night falls in Washington, D.C., where the streets are also being cleaned—by a handful of giant street-sweeping machines driven by a handful of workers.

The difference in method is not the result of a greater knowledge of modern technology in the United States—the Chinese know perfectly well how to build street-sweeping machines. It is a production decision based on costs in the two countries. In China, where wages are relatively low, an army of workers armed with brooms is the least expensive way to produce clean streets. In Washington, where labor costs are high, it makes sense to use more machinery and less labor.

All types of production efforts require choices in the use of factors of production. In this chapter we examine such choices. Should a good or service be produced using relatively more labor and less capital? Or should relatively more capital and less labor be used? What about the use of natural resources?

In this chapter we see why firms make the production choices they do and how their costs affect their choices. We will apply the marginal decision rule to the production process and see how this rule ensures that production is carried out at the lowest cost possible. We examine the nature of production and costs in order to gain a better understanding of supply. We thus shift our focus to **firms**, organizations that produce goods and services. In producing goods and services, firms combine the factors of production—labor, capital, and natural resources—to produce various products.

firms

Organizations that produce goods and services.

Economists assume that firms engage in production in order to earn a profit and that they seek to make this profit as large as possible. That is, economists assume that firms apply the marginal decision rule as they seek to maximize their profits. Whether we consider the operator of a shoe-shine stand at an airport or the firm that produces airplanes, we will find there are basic relationships between the use of factors of production and output levels, and between output levels and costs, that apply to all production. The production choices of firms and their associated costs are at the foundation of supply.

1. PRODUCTION CHOICES AND COSTS: THE SHORT RUN

LEARNING OBJECTIVES

1. **Understand the terms associated with the short-run production function—total product, average product, and marginal product—and explain and illustrate how they are related to each other.**
2. **Explain the concepts of increasing, diminishing, and negative marginal returns and explain the law of diminishing marginal returns.**
3. **Understand the terms associated with costs in the short run—total variable cost, total fixed cost, total cost, average variable cost, average fixed cost, average total cost, and marginal cost—and explain and illustrate how they are related to each other.**
4. **Explain and illustrate how the product and cost curves are related to each other and to determine in what ranges on these curves marginal returns are increasing, diminishing, or negative.**

short run

A planning period over which the managers of a firm must consider one or more of their factors of production as fixed in quantity.

Our analysis of production and cost begins with a period economists call the short run. The **short run** in this microeconomic context is a planning period over which the managers of a firm must consider one or more of their factors of production as fixed in quantity. For example, a restaurant may regard its building as a fixed factor over a period of at least the next year. It would take at least that much time to find a new building or to expand or reduce the size of its present facility. Decisions concerning the operation of the restaurant during the next year must assume the building will remain unchanged. Other factors of production could be changed during the year, but the size of the building must be regarded as a constant.

fixed factor of production

A factor of production whose quantity cannot be changed during a particular period.

variable factor of production

A factor of production whose quantity can be changed during a particular period.

When the quantity of a factor of production cannot be changed during a particular period, it is called a **fixed factor of production**. For the restaurant, its building is a fixed factor of production for at least a year. A factor of production whose quantity can be changed during a particular period is called a **variable factor of production**; factors such as labor and food are examples.

long run

The planning period over which a firm can consider all factors of production as variable.

While the managers of the restaurant are making choices concerning its operation over the next year, they are also planning for longer periods. Over those periods, managers may contemplate alternatives such as modifying the building, building a new facility, or selling the building and leaving the restaurant business. The planning period over which a firm can consider *all* factors of production as variable is called the **long run**.

At any one time, a firm will be making both short-run and long-run choices. The managers may be planning what to do for the next few weeks and for the next few years. Their decisions over the next few weeks are likely to be short-run choices. Decisions that will affect operations over the next few years may be long-run choices, in which managers can consider changing every aspect of their operations. Our analysis in this section focuses on the short run. We examine long-run choices later in this chapter.

1.1 The Short-Run Production Function

production function

The relationship between factors of production and the output of a firm.

A firm uses factors of production to produce a product. The relationship between factors of production and the output of a firm is called a **production function** Our first task is to explore the nature of the production function.

Consider a hypothetical firm, Acme Clothing, a shop that produces jackets. Suppose that Acme has a lease on its building and equipment. During the period of the lease, Acme's capital is its fixed factor of production. Acme's variable factors of production include things such as labor, cloth, and electricity. In the analysis that follows, we shall simplify by assuming that labor is Acme's *only* variable factor of production.

Total, Marginal, and Average Products

Figure 8.1 shows the number of jackets Acme can obtain with varying amounts of labor (in this case, tailors) and its given level of capital. A **total product curve** shows the quantities of output that can be obtained from different amounts of a variable factor of production, assuming other factors of production are fixed.

total product curve

Graph that shows the quantities of output that can be obtained from different amounts of a variable factor of production, assuming other factors of production are fixed.

Notice what happens to the slope of the total product curve in Figure 8.1. Between 0 and 3 units of labor per day, the curve becomes steeper. Between 3 and 7 workers, the curve continues to slope upward, but its slope diminishes. Beyond the seventh tailor, production begins to decline and the curve slopes downward.

We measure the slope of any curve as the vertical change between two points divided by the horizontal change between the same two points. The slope of the total product curve for labor equals the change in output (ΔQ) divided by the change in units of labor (ΔL):

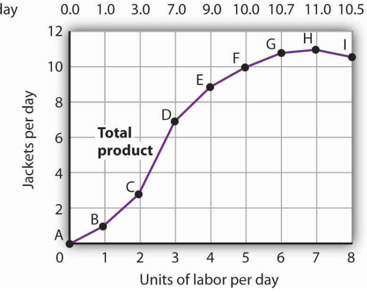
$$\text{Slope of the total product curve} = \Delta Q / \Delta L$$

The slope of a total product curve for any variable factor is a measure of the change in output associated with a change in the amount of the variable factor, with the quantities of all other factors held constant. The amount by which output rises with an additional unit of a variable factor is the **marginal product** of the variable factor. Mathematically, marginal product is the ratio of the change in output to the change in the amount of a variable factor. The **marginal product of labor** (MP_L), for example, is the amount by which output rises with an additional unit of labor. It is thus the ratio of the change in output to the change in the quantity of labor ($\Delta Q/\Delta L$), all other things unchanged. It is measured as the slope of the total product curve for labor.

FIGURE 8.1 Acme Clothing's Total Product Curve

The table gives output levels per day for Acme Clothing Company at various quantities of labor per day, assuming the firm's capital is fixed. These values are then plotted graphically as a total product curve.

Point on graph	A	B	C	D	E	F	G	H	I
Units of labor per day	0	1	2	3	4	5	6	7	8
Jackets per day	0.0	1.0	3.0	7.0	9.0	10.0	10.7	11.0	10.5



EQUATION 8.1

$$MP_L = \Delta Q / \Delta L$$

In addition we can define the **average product** of a variable factor. It is the output per unit of variable factor. The **average product of labor** (AP_L), for example, is the ratio of output to the number of units of labor (Q/L).

EQUATION 8.2

$$AP_L = Q / L$$

marginal product

The amount by which output rises with an additional unit of a variable factor.

marginal product of labor

The amount by which output rises with an additional unit of labor.

average product

The output per unit of variable factor.

average product of labor

The ratio of output to the number of units of labor (Q/L).

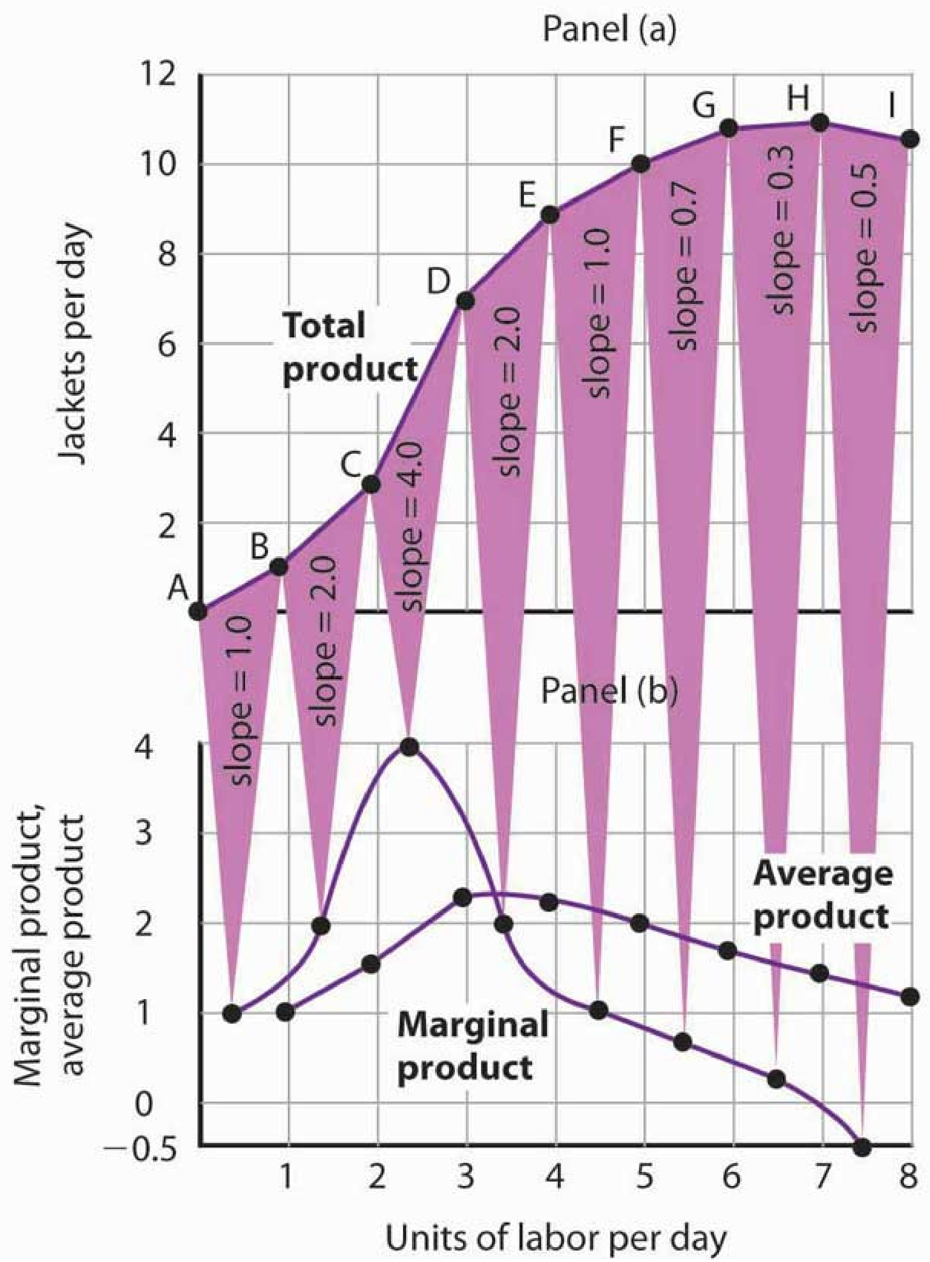
The concept of average product is often used for comparing productivity levels over time or in comparing productivity levels among nations. When you read in the newspaper that productivity is rising or falling, or that productivity in the United States is nine times greater than productivity in China, the report is probably referring to some measure of the average product of labor.

The total product curve in Panel (a) of Figure 8.2 is repeated from Figure 8.1. Panel (b) shows the marginal product and average product curves. Notice that marginal product is the slope of the total product curve, and that marginal product rises as the slope of the total product curve increases, falls as the slope of the total product curve declines, reaches zero when the total product curve achieves its maximum value, and becomes negative as the total product curve slopes downward. As in other parts of this text, marginal values are plotted at the midpoint of each interval. The marginal product of the fifth unit of labor, for example, is plotted between 4 and 5 units of labor. Also notice that the marginal product curve intersects the average product curve at the maximum point on the average product curve. When marginal product is above average product, average product is rising. When marginal product is below average product, average product is falling.

FIGURE 8.2 From Total Product to the Average and Marginal Product of Labor

The first two rows of the table give the values for quantities of labor and total product from Figure 8.1. Marginal product, given in the third row, is the change in output resulting from a one-unit increase in labor. Average product, given in the fourth row, is output per unit of labor. Panel (a) shows the total product curve. The slope of the total product curve is marginal product, which is plotted in Panel (b). Values for marginal product are plotted at the midpoints of the intervals. Average product rises and falls. Where marginal product is above average product, average product rises. Where marginal product is below average product, average product falls. The marginal product curve intersects the average product curve at the maximum point on the average product curve.

Units of labor per day	0	1	2	3	4	5	6	7	8
Jackets per day	0	1.0	3.0	7.0	9.0	10.0	10.7	11.0	10.5
Marginal product		1.0	2.0	4.0	2.0	1.0	0.7	0.3	−0.5
Average product	—	1.0	1.5	2.33	2.25	2.0	1.78	1.57	1.31



As a student you can use your own experience to understand the relationship between marginal and average values. Your grade point average (GPA) represents the average grade you have earned in all your course work so far. When you take an additional course, your grade in that course represents the marginal grade. What happens to your GPA when you get a grade that is higher than your previous average? It rises. What happens to your GPA when you get a grade that is lower than your previous average? It falls. If your GPA is a 3.0 and you earn one more B, your marginal grade equals your GPA and your GPA remains unchanged.

The relationship between average product and marginal product is similar. However, unlike your course grades, which may go up and down willy-nilly, marginal product always rises and then falls, for reasons we will explore shortly. As soon as marginal product falls below average product, the average product curve slopes downward. While marginal product is above average product, whether marginal product is increasing or decreasing, the average product curve slopes upward.

As we have learned, maximizing behavior requires focusing on making decisions at the margin. For this reason, we turn our attention now toward increasing our understanding of marginal product.

Increasing, Diminishing, and Negative Marginal Returns

Adding the first worker increases Acme's output from 0 to 1 jacket per day. The second tailor adds 2 jackets to total output; the third adds 4. The marginal product goes up because when there are more workers, each one can specialize to a degree. One worker might cut the cloth, another might sew the seams, and another might sew the buttonholes. Their increasing marginal products are reflected by the increasing slope of the total product curve over the first 3 units of labor and by the upward slope of the marginal product curve over the same range. The range over which marginal products are increasing is called the range of **increasing marginal returns**. Increasing marginal returns exist in the context of a total product curve for labor, so we are holding the quantities of other factors constant. Increasing marginal returns may occur for any variable factor.

increasing marginal returns

The range over which each additional unit of a variable factor adds more to total output than the previous unit.

The fourth worker adds less to total output than the third; the marginal product of the fourth worker is 2 jackets. The data in Figure 8.2 show that marginal product continues to decline after the fourth worker as more and more workers are hired. The additional workers allow even greater opportunities for specialization, but because they are operating with a fixed amount of capital, each new worker adds less to total output. The fifth tailor adds only a single jacket to total output. When each additional unit of a variable factor adds less to total output, the firm is experiencing **diminishing marginal returns**. Over the range of diminishing marginal returns, the marginal product of the variable factor is positive but falling. Once again, we assume that the quantities of all other factors of production are fixed. Diminishing marginal returns may occur for any variable factor. Panel (b) shows that Acme experiences diminishing marginal returns between the third and seventh workers, or between 7 and 11 jackets per day.

diminishing marginal returns

The range over which each additional unit of a variable factor adds less to total output than the previous unit.

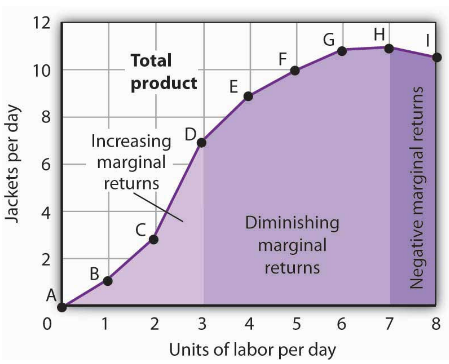
After the seventh unit of labor, Acme's fixed plant becomes so crowded that adding another worker actually reduces output. When additional units of a variable factor reduce total output, given constant quantities of all other factors, the company experiences **negative marginal returns**. Now the total product curve is downward sloping, and the marginal product curve falls below zero. Figure 8.3 shows the ranges of increasing, diminishing, and negative marginal returns. Clearly, a firm will never intentionally add so much of a variable factor of production that it enters a range of negative marginal returns.

negative marginal returns

The range over which additional units of a variable factor reduce total output, given constant quantities of all other factors.

FIGURE 8.3 Increasing Marginal Returns, Diminishing Marginal Returns, and Negative Marginal Returns

This graph shows Acme's total product curve from Figure 8.1 with the ranges of increasing marginal returns, diminishing marginal returns, and negative marginal returns marked. Acme experiences increasing marginal returns between 0 and 3 units of labor per day, diminishing marginal returns between 3 and 7 units of labor per day, and negative marginal returns beyond the 7th unit of labor.



The idea that the marginal product of a variable factor declines over some range is important enough, and general enough, that economists state it as a law. The **law of diminishing marginal returns** holds that the marginal product of any variable factor of production will eventually decline, assuming the quantities of other factors of production are unchanged.

Heads Up!

It is easy to confuse the concept of diminishing marginal returns with the idea of negative marginal returns. To say a firm is experiencing diminishing marginal returns is not to say its output is falling. Diminishing marginal returns mean that the marginal product of a variable factor is declining. Output is still increasing as the variable factor is increased, but it is increasing by smaller and smaller amounts. As we saw in Figure 8.2 and Figure 8.3, the range of diminishing marginal returns was between the third and seventh workers; over this range of workers, output rose from 7 to 11 jackets. Negative marginal returns started after the seventh worker.

To see the logic of the law of diminishing marginal returns, imagine a case in which it does not hold. Say that you have a small plot of land for a vegetable garden, 10 feet by 10 feet in size. The plot itself is a fixed factor in the production of vegetables. Suppose you are able to hold constant all other factors—water, sunshine, temperature, fertilizer, and seed—and vary the amount of labor devoted to the garden. How much food could the garden produce? Suppose the marginal product of labor kept increasing or was constant. Then you could grow an *unlimited* quantity of food on your small plot—enough to feed the entire world! You could add an unlimited number of workers to your plot and still increase output at a constant or increasing rate. If you did not get enough output with, say, 500 workers, you could use 5 million; the five-millionth worker would add at least as much to total output as the first. If diminishing marginal returns to labor did not occur, the total product curve would slope upward at a constant or increasing rate.

law of diminishing marginal returns

The marginal product of any variable factor of production will eventually decline, assuming the quantities of other factors of production are unchanged.

The shape of the total product curve and the shape of the resulting marginal product curve drawn in Figure 8.2 are typical of *any* firm for the short run. Given its fixed factors of production, increasing the use of a variable factor will generate increasing marginal returns at first; the total product curve for the variable factor becomes steeper and the marginal product rises. The opportunity to gain from increased specialization in the use of the variable factor accounts for this range of increasing marginal returns. Eventually, though, diminishing returns will set in. The total product curve will become flatter, and the marginal product curve will fall.

1.2 Costs in the Short Run

A firm's costs of production depend on the quantities and prices of its factors of production. Because we expect a firm's output to vary with the firm's use of labor in a specific way, we can also expect the firm's costs to vary with its output in a specific way. We shall put our information about Acme's product curves to work to discover how a firm's costs vary with its level of output.

We distinguish between the costs associated with the use of variable factors of production, which are called **variable costs**, and the costs associated with the use of fixed factors of production, which are called **fixed costs**. For most firms, variable costs includes costs for raw materials, salaries of production workers, and utilities. The salaries of top management may be fixed costs; any charges set by contract over a period of time, such as Acme's one-year lease on its building and equipment, are likely to be fixed costs. A term commonly used for fixed costs is *overhead*. Notice that fixed costs exist only in the short run. In the long run, the quantities of all factors of production are variable, so that all long-run costs are variable.

variable costs

The costs associated with the use of variable factors of production.

fixed costs

The costs associated with the use of fixed factors of production.

Total variable cost (TVC) is cost that varies with the level of output. **Total fixed cost (TFC)** is cost that does not vary with output. **Total cost (TC)** is the sum of total variable cost and total fixed cost:

total variable cost

Cost that varies with the level of output.

total fixed cost

Cost that does not vary with output.

total cost

The sum of total variable cost and total fixed cost.

EQUATION 8.3

$$TVC + TFC = TC$$

From Total Production to Total Cost

Next we illustrate the relationship between Acme's total product curve and its total costs. Acme can vary the quantity of labor it uses each day, so the cost of this labor is a variable cost. We assume capital is a fixed factor of production in the short run, so its cost is a fixed cost.

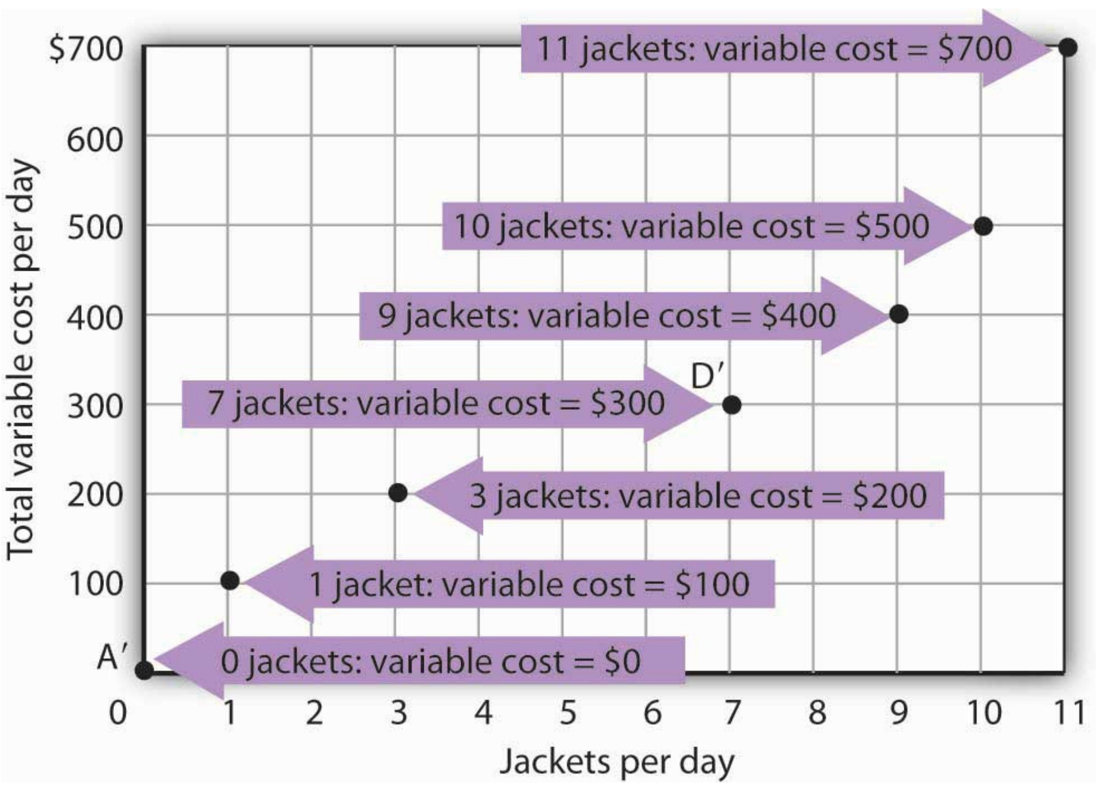
Suppose that Acme pays a wage of $100 per worker per day. If labor is the only variable factor, Acme's total variable costs per day amount to $100 times the number of workers it employs. We can use the information given by the total product curve, together with the wage, to compute Acme's total variable costs.

We know from Figure 8.1 that Acme requires 1 worker working 1 day to produce 1 jacket. The total variable cost of a jacket thus equals $100. Three units of labor produce 7 jackets per day; the total variable cost of 7 jackets equals $300. Figure 8.4 shows Acme's total variable costs for producing each of the output levels given in Figure 8.1

Figure 8.4 gives us costs for several quantities of jackets, but we need a bit more detail. We know, for example, that 7 jackets have a total variable cost of $300. What is the total variable cost of 6 jackets?

FIGURE 8.4 Computing Variable Costs

The points shown give the variable costs of producing the quantities of jackets given in the total product curve in Figure 8.1 and Figure 8.2. Suppose Acme's workers earn $100 per day. If Acme produces 0 jackets, it will use no labor—its variable cost thus equals $0 (Point A′). Producing 7 jackets requires 3 units of labor; Acme's variable cost equals $300 (Point D′).

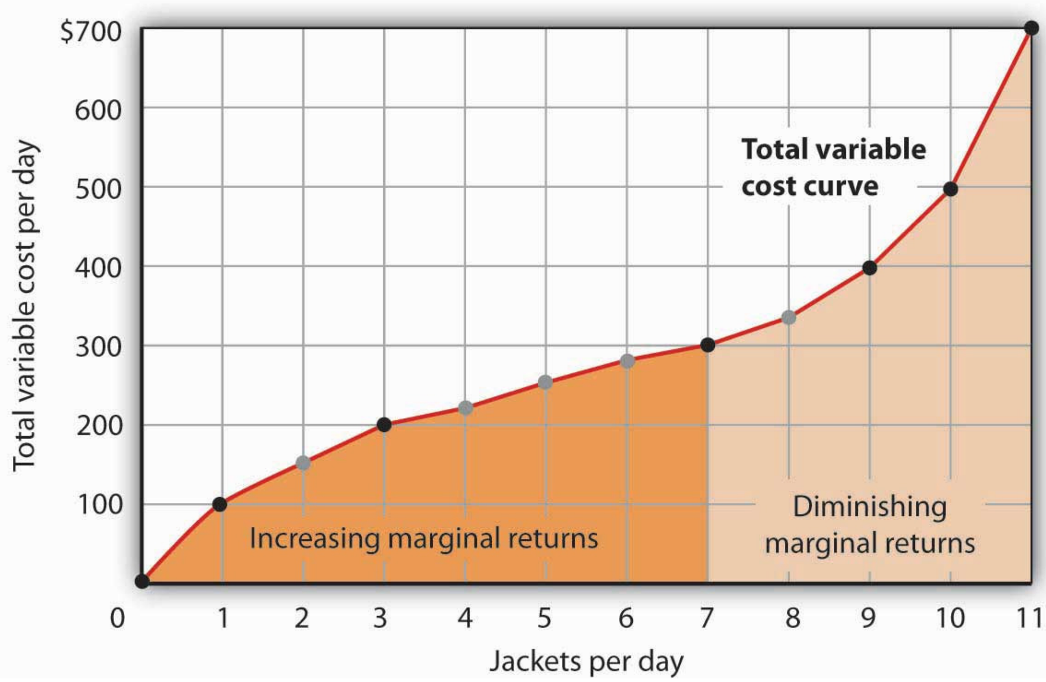


We can estimate total variable costs for other quantities of jackets by inspecting the total product curve in Figure 8.1. Reading over from a quantity of 6 jackets to the total product curve and then down suggests that the Acme needs about 2.8 units of labor to produce 6 jackets per day. Acme needs 2 full-time and 1 part-time tailors to produce 6 jackets. Figure 8.5 gives the precise total variable costs for quantities of jackets ranging from 0 to 11 per day. The numbers in boldface type are taken from Figure 8.4; the other numbers are estimates we have assigned to produce a total variable cost curve that is consistent with our total product curve. You should, however, be certain that you understand how the numbers in boldface type were found.

FIGURE 8.5 The Total Variable Cost Curve

Total variable costs for output levels shown in Acme's total product curve were shown in Figure 8.4. To complete the total variable cost curve, we need to know the variable cost for each level of output from 0 to 11 jackets per day. The variable costs and quantities of labor given in Figure 8.4 are shown in boldface in the table here and with black dots in the graph. The remaining values were estimated from the total product curve in Figure 8.1 and Figure 8.2. For example, producing 6 jackets requires 2.8 workers, for a variable cost of $280.

Quantity/day	0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0
Labor/day	**0**	**1.00**	1.63	**2.00**	2.33	2.58	2.80	**3.00**	3.38	**4.00**	**5.00**	**7.00**
Total variable cost	**$0**	**$100**	$163	**$200**	$233	$258	$280	**$300**	$338	**$400**	**$500**	**$700**

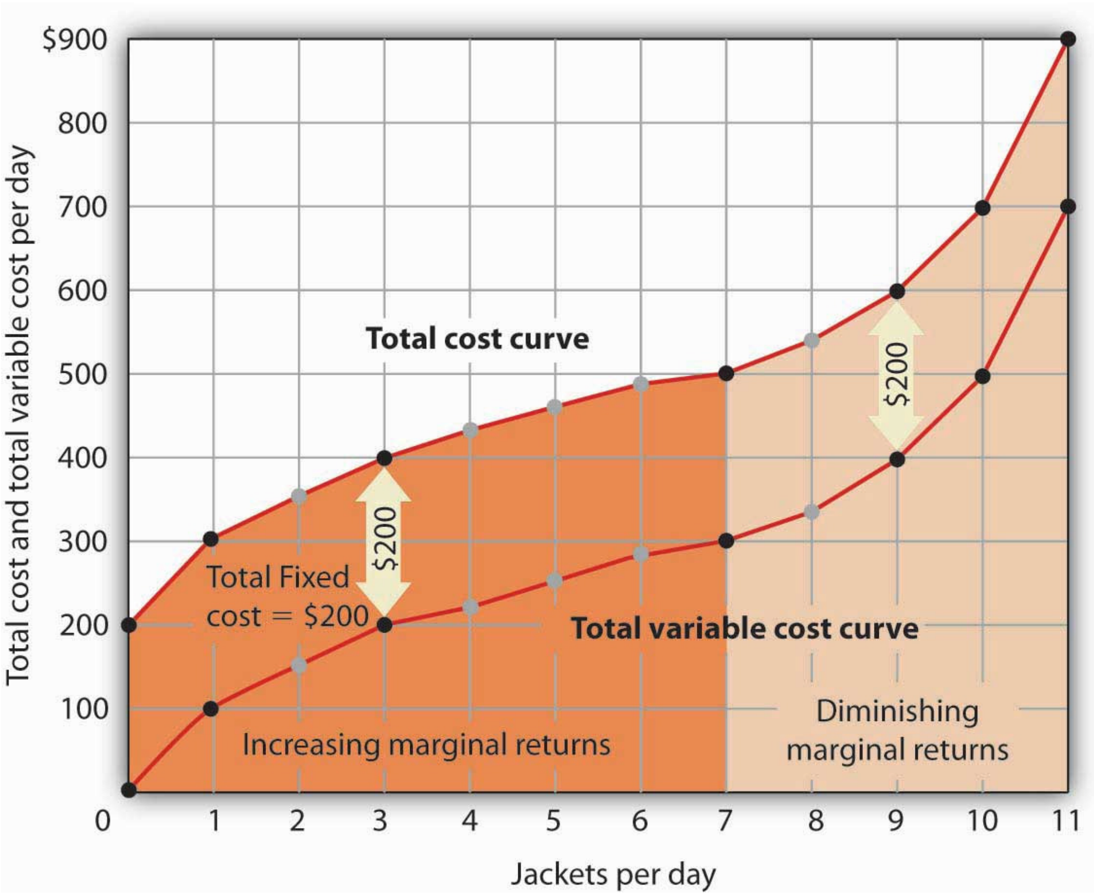


Suppose Acme's present plant, including the building and equipment, is the equivalent of 20 units of capital. Acme has signed a long-term lease for these 20 units of capital at a cost of $200 per day. In the short run, Acme cannot increase or decrease its quantity of capital—it must pay the $200 per day no matter what it does. Even if the firm cuts production to zero, it must still pay $200 per day in the short run.

Acme's total cost is its total fixed cost of $200 plus its total variable cost. We add $200 to the total variable cost curve in Figure 8.5 to get the total cost curve shown in Figure 8.6.

FIGURE 8.6 From Variable Cost to Total Cost

We add total fixed cost to the total variable cost to obtain total cost. In this case, Acme's total fixed cost equals $200 per day.



Notice something important about the shapes of the total cost and total variable cost curves in Figure 8.6. The total cost curve, for example, starts at $200 when Acme produces 0 jackets—that is its total fixed cost. The curve rises, but at a decreasing rate, up to the seventh jacket. Beyond the seventh jacket, the curve becomes steeper and steeper. The slope of the total variable cost curve behaves in precisely the same way.

Recall that Acme experienced increasing marginal returns to labor for the first three units of labor—or the first seven jackets. Up to the third worker, each additional worker added more and more to Acme's output. Over the range of increasing marginal returns, each additional jacket requires less and less additional labor. The first jacket required one tailor; the second required the addition of only a part-time tailor; the third required only that Acme boost that part-time tailor's hours to a full day. Up to the seventh jacket, each additional jacket requires less and less additional labor, and thus costs rise at a decreasing rate; the total cost and total variable cost curves become flatter over the range of increasing marginal returns.

Acme experiences diminishing marginal returns beyond the third unit of labor—or the seventh jacket. Notice that the total cost and total variable cost curves become steeper and steeper beyond this level of output. In the range of diminishing marginal returns, each additional unit of a factor adds less and less to total output. That means each additional unit of output requires larger and larger increases in the variable factor, and larger and larger increases in costs.

Marginal and Average Costs

average total cost

Total cost divided by quantity; it is the firms total cost per unit of output.

Marginal and average cost curves, which will play an important role in the analysis of the firm, can be derived from the total cost curve. *Marginal cost* shows the additional cost of each additional unit of output a firm produces. This is a specific application of the general concept of marginal cost presented earlier. Given the marginal decision rule's focus on evaluating choices at the margin, the marginal cost curve takes on enormous importance in the analysis of a firm's choices. The second curve we shall derive shows the firm's average total cost at each level of output. **Average total cost (ATC)** is total cost divided by quantity; it is the firm's total cost per unit of output:

EQUATION 8.4

$$ATC = TC / Q$$

average variable cost

Total variable cost divided by quantity; it is the firm's total variable cost per unit of output.

We shall also discuss **average variable costs** *(AVC)*, which is the firm's variable cost per unit of output; it is total variable cost divided by quantity:

EQUATION 8.5

$$AVC = TVC / Q$$

average fixed cost

Total fixed cost divided by quantity.

We are still assessing the choices facing the firm in the short run, so we assume that at least one factor of production is fixed. Finally, we will discuss **average fixed cost** *(AFC)*, which is total fixed cost divided by quantity:

EQUATION 8.6

$$AFC = TFC / Q$$

Marginal cost *(MC)* is the amount by which total cost rises with an additional unit of output. It is the ratio of the change in total cost to the change in the quantity of output:

EQUATION 8.7

$$MC = \Delta TC / \Delta Q$$

It equals the slope of the total cost curve. Figure 8.7 shows the same total cost curve that was presented in Figure 8.6. This time the slopes of the total cost curve are shown; these slopes equal the marginal cost of each additional unit of output. For example, increasing output from 6 to 7 units ($\Delta Q = 1$) increases total cost from \$480 to \$500 ($\Delta TC = \$20$). The seventh unit thus has a marginal cost of \$20 ($\Delta TC / \Delta Q = \$20 / 1 = \$20$). Marginal cost falls over the range of increasing marginal returns and rises over the range of diminishing marginal returns.

Heads Up!

Notice that the various cost curves are drawn with the quantity of output on the horizontal axis. The various product curves are drawn with quantity of a factor of production on the horizontal axis. The reason is that the two sets of curves measure different relationships. Product curves show the relationship between output and the quantity of a factor; they therefore have the factor quantity on the horizontal axis. Cost curves show how costs vary with output and thus have output on the horizontal axis.

FIGURE 8.7 **Total Cost and Marginal Cost**

Marginal cost in Panel (b) is the slope of the total cost curve in Panel (a).

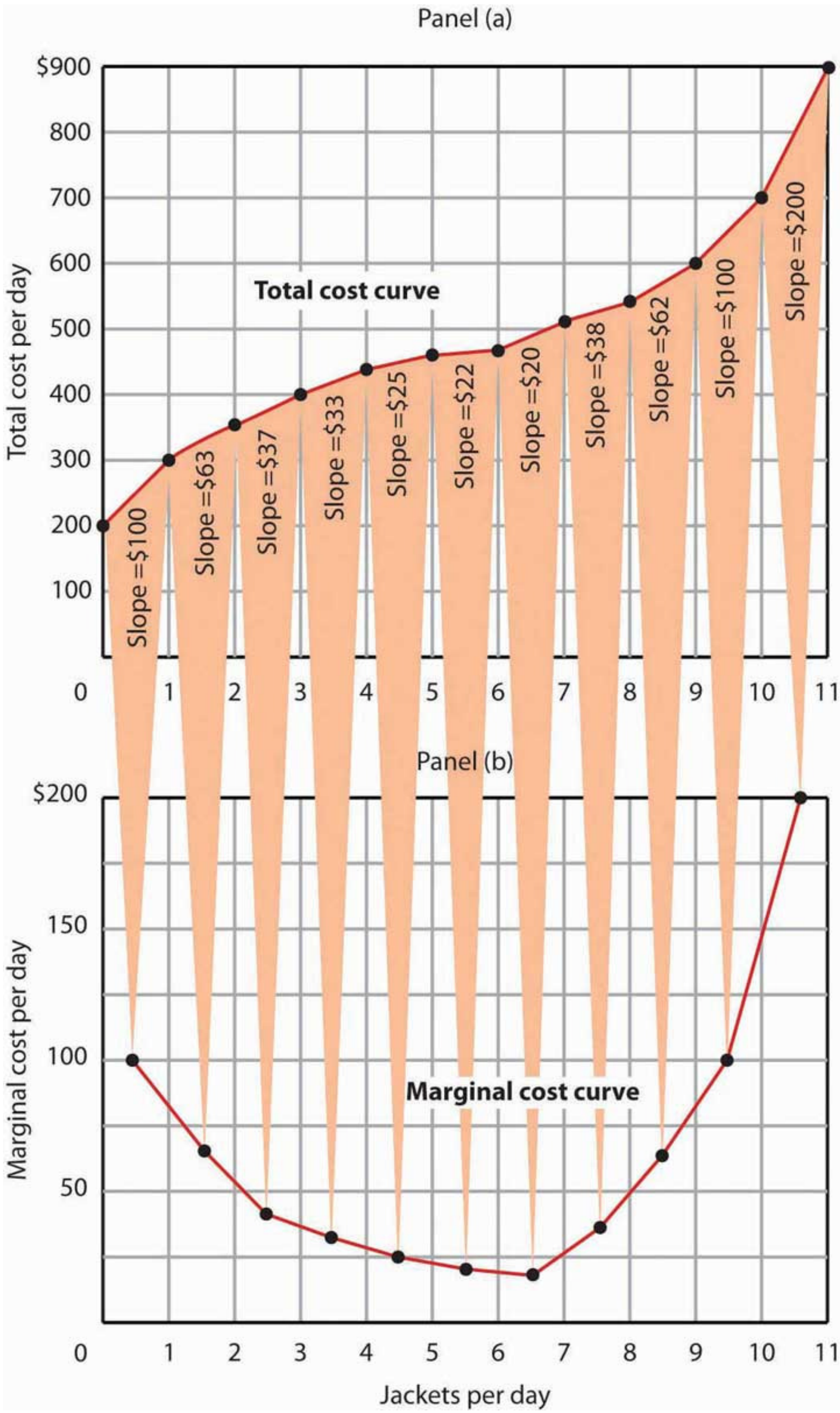


Figure 8.8 shows the computation of Acme's short-run average total cost, average variable cost, and average fixed cost and graphs of these values. Notice that the curves for short-run average total cost and average variable cost fall, then rise. We say that these cost curves are U-shaped. Average fixed cost keeps falling as output increases. This is because the fixed costs are spread out more and more as

output expands; by definition, they do not vary as labor is added. Since average total cost (*ATC*) is the sum of average variable cost (*AVC*) and average fixed cost (*AFC*), i.e.,

EQUATION 8.8

$$AVC + AFC = ATC$$

the distance between the *ATC* and *AVC* curves keeps getting smaller and smaller as the firm spreads its overhead costs over more and more output.

FIGURE 8.8 Marginal Cost, Average Fixed Cost, Average Variable Cost, and Average Total Cost in the Short Run

Total cost figures for Acme Clothing are taken from Figure 8.7. The other values are derived from these. Average total cost (*ATC*) equals total cost divided by quantity produced; it also equals the sum of the average fixed cost (*AFC*) and average variable cost (*AVC*) (exceptions in table are due to rounding to the nearest dollar); average variable cost is variable cost divided by quantity produced. The marginal cost (*MC*) curve (from Figure 8.7) intersects the *ATC* and *AVC* curves at the lowest points on both curves. The *AFC* curve falls as quantity increases.

Quantity/day	0	1	2	3	4	5	6	7	8	9	10	11
Total cost	\$200	\$300	\$363	\$400	\$433	\$458	\$480	\$500	\$538	\$600	\$700	\$900
AFC		\$200	\$100	\$67	\$50	\$40	\$33	\$29	\$25	\$22	\$20	\$18
AVC		\$100	\$82	\$67	\$58	\$52	\$47	\$43	\$42	\$44	\$50	\$64
ATC		\$300	\$182	\$133	\$108	\$92	\$80	\$71	\$67	\$67	\$70	\$82
MC		\$100	\$63	\$37	\$33	\$25	\$22	\$20	\$38	\$62	\$100	\$200

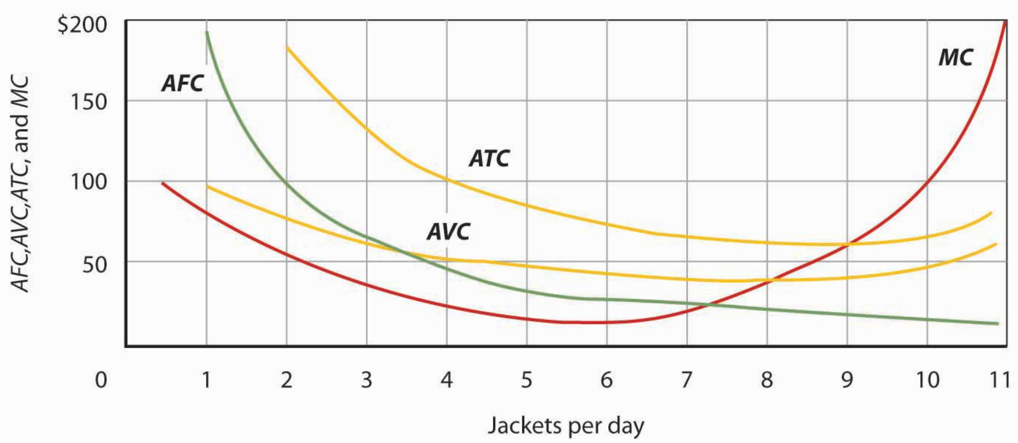
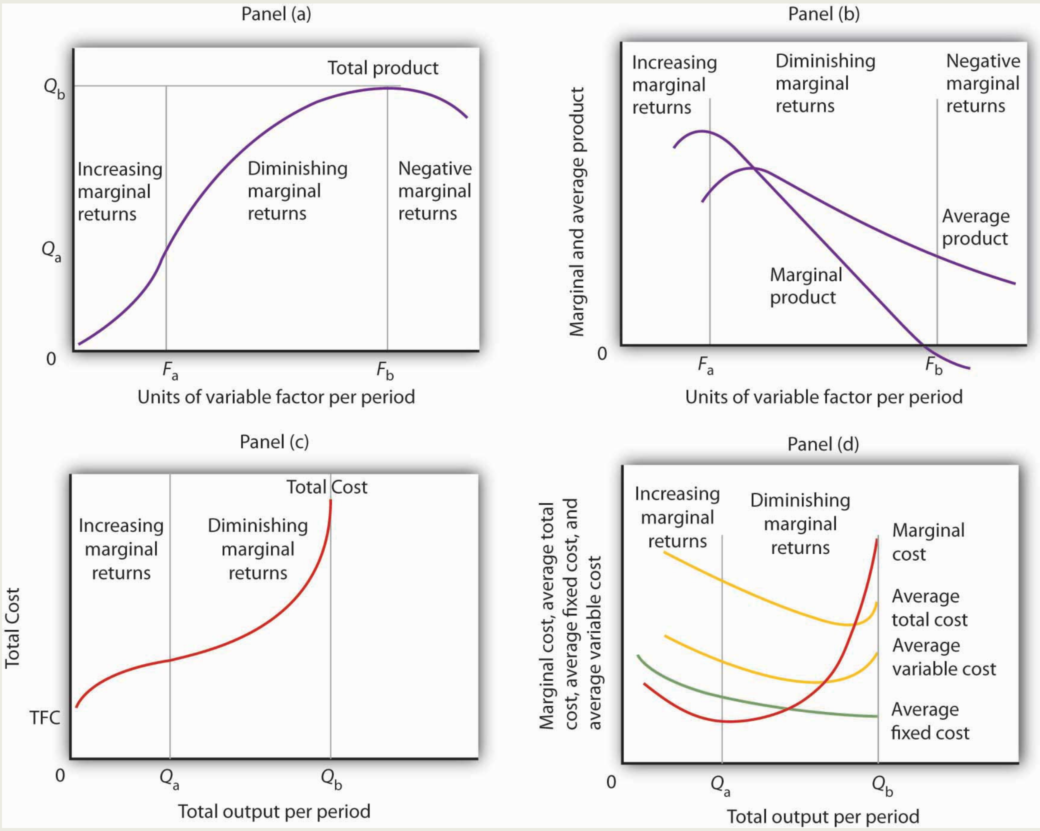


Figure 8.8 includes the marginal cost data and the marginal cost curve from Figure 8.7. The marginal cost curve intersects the average total cost and average variable cost curves at their lowest points. When marginal cost is below average total cost or average variable cost, the average total and average variable cost curves slope downward. When marginal cost is greater than short-run average total cost or average variable cost, these average cost curves slope upward. The logic behind the relationship between marginal cost and average total and variable costs is the same as it is for the relationship between marginal product and average product.

We turn next in this chapter to an examination of production and cost in the long run, a planning period in which the firm can consider changing the quantities of any or all factors.

KEY TAKEAWAYS



- In Panel (a), the total product curve for a variable factor in the short run shows that the firm experiences increasing marginal returns from zero to F_a units of the variable factor (zero to Q_a units of output), diminishing marginal returns from F_a to F_b (Q_a to Q_b units of output), and negative marginal returns beyond F_b units of the variable factor.
- Panel (b) shows that marginal product rises over the range of increasing marginal returns, falls over the range of diminishing marginal returns, and becomes negative over the range of negative marginal returns. Average product rises when marginal product is above it and falls when marginal product is below it.
- In Panel (c), total cost rises at a decreasing rate over the range of output from zero to Q_a This was the range of output that was shown in Panel (a) to exhibit increasing marginal returns. Beyond Q_a, the range of diminishing marginal returns, total cost rises at an increasing rate. The total cost at zero units of output (shown as the intercept on the vertical axis) is total fixed cost.
- Panel (d) shows that marginal cost falls over the range of increasing marginal returns, then rises over the range of diminishing marginal returns. The marginal cost curve intersects the average total cost and average variable cost curves at their lowest points. Average fixed cost falls as output increases. Note that average total cost equals average variable cost plus average fixed cost.
- Assuming labor is the variable factor of production, the following definitions and relations describe production and cost in the short run:

$$MP_L = \Delta Q / \Delta L$$

$$AP_L = Q / L$$

$$TVC + TFC = TC$$

$$ATC = TC / Q$$

$$AVC = TVC / Q$$

$$AFC = TFC / Q$$

$$MC = \Delta TC / \Delta Q$$

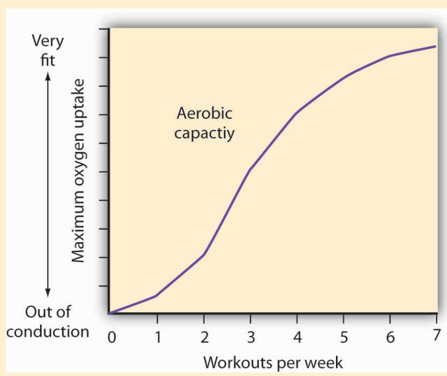
TRY IT!

1. Suppose Acme gets some new equipment for producing jackets. The table below gives its new production function. Compute marginal product and average product and fill in the bottom two rows of the table. Referring to Figure 8.2, draw a graph showing Acme's new total product curve. On a second graph, below the one showing the total product curve you drew, sketch the marginal and average product curves. Remember to plot marginal product at the midpoint between each input level. On both graphs, shade the regions where Acme experiences increasing marginal returns, diminishing marginal returns, and negative marginal returns.

Units of labor per day	0	1	2	3	4	5	6	7	8
Jackets per day	0	2	5.5	9.5	12	14	15	15.5	15
Marginal product	___	___	___	___	___	___	___	___	
Average product	___	___	___	___	___	___	___	___	___

2. Draw the points showing total variable cost at daily outputs of 0, 1, 3, 7, 9, 10, and 11 jackets per day when Acme faced a wage of $100 per day. (Use Figure 8.5 as a model.) Sketch the total variable cost curve as shown in Figure 8.4. Now suppose that the wage rises to $125 per day. On the same graph, show the new points and sketch the new total variable cost curve. Explain what has happened. What will happen to Acme's marginal cost curve? Its average total, average variable, and average fixed cost curves? Explain.

Case in Point: The Production of Fitness



How much should an athlete train?

Sports physiologists often measure the "total product" of training as the increase in an athlete's aerobic capacity—the capacity to absorb oxygen into the bloodstream. An athlete can be thought of as producing aerobic capacity using a fixed factor (his or her natural capacity) and a variable input (exercise). The chart shows how this aerobic capacity varies with the number of workouts per week. The curve has a shape very much like a total product curve—which, after all, is precisely what it is.

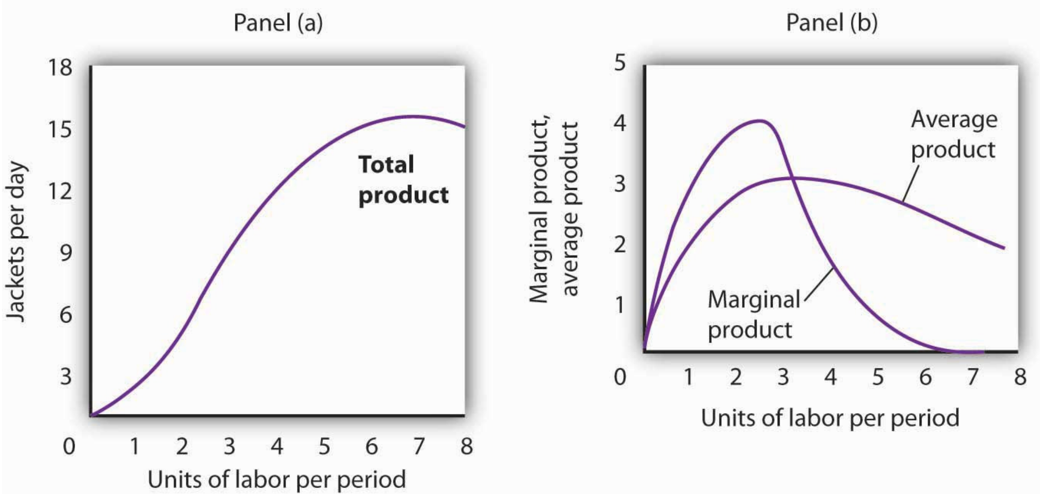
The data suggest that an athlete experiences increasing marginal returns from exercise for the first three days of training each week; indeed, over half the total gain in aerobic capacity possible is achieved. A person can become even more fit by exercising more, but the gains become smaller with each added day of training. The law of diminishing marginal returns applies to training.

The increase in fitness that results from the sixth and seventh workouts each week is small. Studies also show that the costs of daily training, in terms of increased likelihood of injury, are high. Many trainers and coaches now recommend that athletes—at all levels of competition—take a day or two off each week.

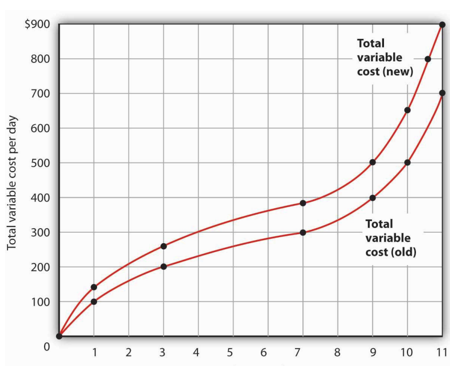
Source: Jeff Galloway, Galloway's Book on Running (Bolinas, CA: Shelter Publications, 2002), p. 56.

ANSWERS TO TRY IT! PROBLEMS

Units of labor per day	0	1	2	3	4	5	6	7	8
Jackets per day	0	2	5.5	9.5	12	14	15	15.5	15
Marginal product		2	3.5	4.0	2.5	2.0	1.0	0.5	−0.5
Average product		2	2.75	3.17	3	2.8	2.5	2.21	1.88



1. The increased wage will shift the total variable cost curve upward; the old and new points and the corresponding curves are shown at the right.
2. The total variable cost curve has shifted upward because the cost of labor, Acme's variable factor, has increased. The marginal cost curve shows the additional cost of each additional unit of output a firm produces. Because an increase in output requires more labor, and because labor now costs more, the marginal cost curve will shift upward. The increase in total variable cost will increase total cost; average total and average variable costs will rise as well. Average fixed cost will not change.



2. PRODUCTION CHOICES AND COSTS: THE LONG RUN

LEARNING OBJECTIVES

1. **Apply the marginal decision rule to explain how a firm chooses its mix of factors of production in the long run.**
2. **Define the long-run average cost curve and explain how it relates to economies and diseconomies or scale.**

In a long-run planning perspective, a firm can consider changing the quantities of *all* its factors of production. That gives the firm opportunities it does not have in the short run. First, the firm can select the mix of factors it wishes to use. Should it choose a production process with lots of labor and not much capital, like the street sweepers in China? Or should it select a process that uses a great deal of capital and relatively little labor, like street sweepers in the United States? The second thing the firm can select is the scale (or overall size) of its operations. In the short run, a firm can increase output only by increasing its use of a variable factor. But in the long run, all factors are variable, so the firm can expand the use of all of its factors of production. The question facing the firm in the long run is: How much of an expansion or contraction in the scale of its operations should it undertake? Alternatively, it could choose to go out of business.

In its long-run planning, the firm not only regards all factors as variable, but it regards all costs as variable as well. There are no fixed costs in the long run. Because all costs are variable, the structure of costs in the long run differs somewhat from what we saw in the short run.

2.1 Choosing the Factor Mix

How shall a firm decide what mix of capital, labor, and other factors to use? We can apply the marginal decision rule to answer this question.

Suppose a firm uses capital and labor to produce a particular good. It must determine how to produce the good and the quantity it should produce. We address the question of how much the firm should produce in subsequent chapters, but certainly the firm will want to produce whatever quantity it chooses at as low a cost as possible. Another way of putting that goal is to say that the firm seeks the maximum output possible at every level of total cost.

At any level of total cost, the firm can vary its factor mix. It could, for example, substitute labor for capital in a way that leaves its total cost unchanged. In terms of the marginal decision rule, we can think of the firm as considering whether to spend an additional $1 on one factor, hence $1 less on another. The marginal decision rule says that a firm will shift spending among factors as long as the marginal benefit of such a shift exceeds the marginal cost.

What is the marginal benefit, say, of an additional $1 spent on capital? An additional unit of capital produces the marginal product of capital. To determine the marginal benefit of $1 spent on capital, we divide capital's marginal product by its price: MP_K/P_K. The price of capital is the "rent" paid for the use of a unit of capital for a given period. If the firm already owns the capital, then this rent is an opportunity cost; it represents the return the firm could get by renting the capital to another user or by selling it and earning interest on the money thus gained.

If capital and labor are the only factors, then spending an additional $1 on capital while holding total cost constant means taking $1 out of labor. The cost of that action will be the output lost from cutting back $1 worth of labor. That cost equals the ratio of the marginal product of labor to the price of labor, MP_L/P_L, where the price of labor is the wage.

Suppose that a firm's marginal product of labor is 15 and the price of labor is $5 per unit; the firm gains 3 units of output by spending an additional $1 on labor. Suppose further that the marginal product of capital is 50 and the price of capital is $50 per unit, so the firm would lose 1 unit of output by spending $1 less on capital.

$$\frac{MP_L}{P_L} > \frac{MP_K}{P_K}$$

$$\frac{15}{5} > \frac{50}{50}$$

The firm achieves a net gain of 2 units of output, without any change in cost, by transferring $1 from capital to labor. It will continue to transfer funds from capital to labor as long as it gains more output from the additional labor than it loses in output by reducing capital. As the firm shifts spending in this fashion, however, the marginal product of labor will fall and the marginal product of capital will rise. At some point, the ratios of marginal product to price will be equal for the two factors. At this point, the firm will obtain the maximum output possible for a given total cost:

EQUATION 8.9

$$\frac{MP_L}{P_L} = \frac{MP_K}{P_K}$$

Suppose that a firm that uses capital and labor is satisfying Equation 8.9 when suddenly the price of labor rises. At the current usage levels of the factors, a higher price of labor (P_L') lowers the ratio of the marginal product of labor to the price of labor:

$$\frac{MP_L}{P_L'} < \frac{MP_K}{P_K}$$

The firm will shift funds out of labor and into capital. It will continue to shift from labor to capital until the ratios of marginal product to price are equal for the two factors. In general, a profit-maximizing firm will seek a combination of factors such that

EQUATION 8.10

$$\frac{MP_1}{P_1} = \frac{MP_2}{P_2} = \ldots = \frac{MP_n}{P_n}$$

When a firm satisfies the condition given in Equation 8.10 for efficient use, it produces the greatest possible output for a given cost. To put it another way, the firm achieves the lowest possible cost for a given level of output.

As the price of labor rises, the firm will shift to a factor mix that uses relatively more capital and relatively less labor. As a firm increases its ratio of capital to labor, we say it is becoming more **capital intensive**. A lower price for labor will lead the firm to use relatively more labor and less capital, reducing its ratio of capital to labor. As a firm reduces its ratio of capital to labor, we say it is becoming more **labor intensive**. The notions of labor-intensive and capital-intensive production are purely relative; they imply only that a firm has a higher or lower ratio of capital to labor.

capital intensive

Situation in which a firm has a high ratio of capital to labor.

labor intensive

Situation in which a firm has a low ratio of labor to capital.

Sometimes economists speak of labor-intensive versus capital-intensive countries in the same manner. One implication of the marginal decision rule for factor use is that firms in countries where labor is relatively expensive, such as the United States, will use capital-intensive production methods. Less developed countries, where labor is relatively cheap, will use labor-intensive methods.

Now that we understand how to apply the marginal decision rule to the problem of choosing the mix of factors, we can answer the question that began this chapter: Why does the United States employ a capital-intensive production process to clean streets while China chooses a labor-intensive process? Given that the same technology—know-how—is available, both countries could, after all, use the same production process. Suppose for a moment that the relative prices of labor and capital are the same in China and the United States. In that case, China and the United States can be expected to use the same method to clean streets. But the price of labor relative to the price of capital is, in fact, far lower in China than in the United States. A lower relative price for labor increases the ratio of the marginal product of labor to its price, making it efficient to substitute labor for capital. China thus finds it cheaper to clean streets with lots of people using brooms, while the United States finds it efficient to clean streets with large machines and relatively less labor.

Maquiladoras, plants in Mexico where processing is done using low-cost workers and labor-intensive methods, allow some U.S. firms to have it both ways. They complete part of the production process in the United States, using capital-intensive methods. They then ship the unfinished goods to *maquiladoras*. For example, many U.S. clothing manufacturers produce cloth at U.S. plants on large high-speed looms. They then ship the cloth to Mexico, where it is fashioned into clothing by workers using sewing machines. Another example is plastic injection molding, which requires highly skilled labor and is made in the U.S. The parts are molded in Texas border towns and are then shipped to *maquiladoras* and used in cars and computers. The resulting items are shipped back to the United States, labeled "Assembled in Mexico from U.S. materials." Overall maquiladoras import 97% of the components they use, of which 80 to 85% come from the U.S.

The *maquiladoras* have been a boon to workers in Mexico, who enjoy a higher demand for their services and receive higher wages as a result. The system also benefits the U.S. firms that participate and U.S. consumers who obtain less expensive goods than they would otherwise. It works because different factor prices imply different mixes of labor and capital. Companies are able to carry out the capital-intensive side of the production process in the United States and the labor-intensive side in Mexico. [1]

2.2 Costs in the Long Run

As in the short run, costs in the long run depend on the firm's level of output, the costs of factors, and the quantities of factors needed for each level of output. The chief difference between long- and short-run costs is there are no fixed factors in the long run. There are thus no fixed costs. All costs are variable, so we do not distinguish between total variable cost and total cost in the long run: total cost *is* total variable cost.

long-run average cost curve

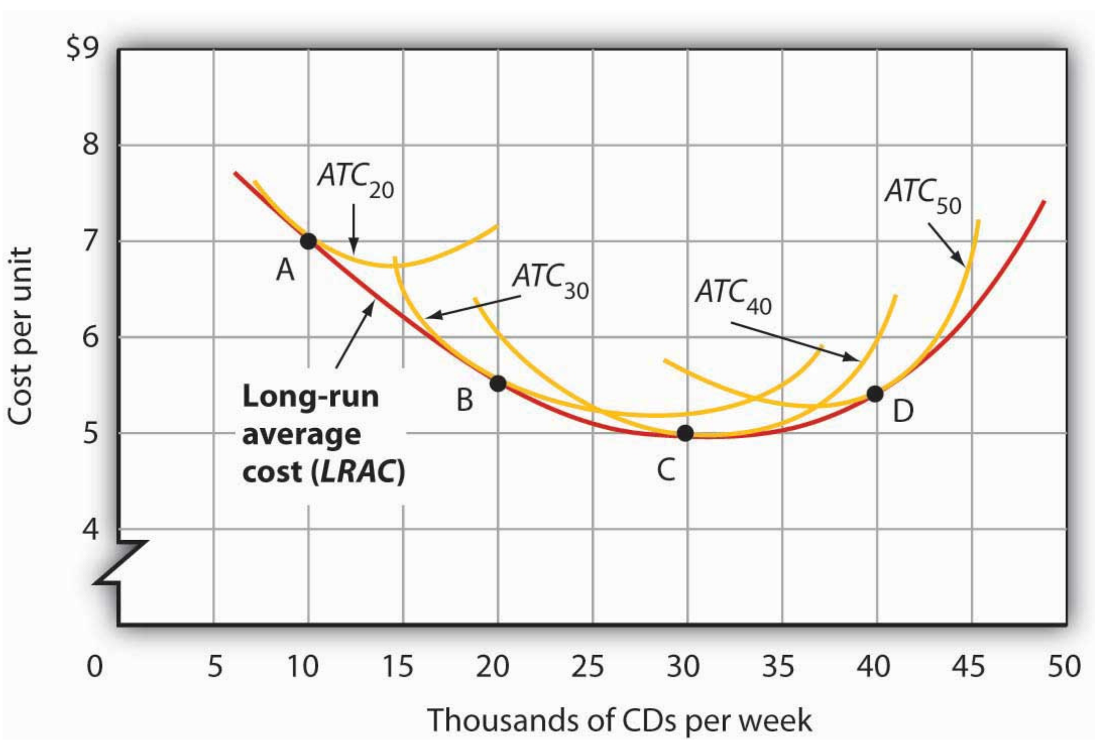
Graph showing the firms lowest cost per unit at each level of output, assuming that all factors of production are variable.

The long-run average cost (LRAC) curve shows the firm's lowest cost per unit at each level of output, assuming that all factors of production are variable. The *LRAC* curve assumes that the firm has chosen the optimal factor mix, as described in the previous section, for producing any level of output. The costs it shows are therefore the lowest costs possible for each level of output. It is important to note, however, that this does not mean that the minimum points of each short-run ATC curves lie on the *LRAC* curve. This critical point is explained in the next paragraph and expanded upon even further in the next section.

Figure 8.14 shows how a firm's *LRAC* curve is derived. Suppose Lifetime Disc Co. produces compact discs (CDs) using capital and labor. We have already seen how a firm's average total cost curve can be drawn in the short run for a given quantity of a particular factor of production, such as capital. In the short run, Lifetime Disc might be limited to operating with a given amount of capital; it would face one of the short-run average total cost curves shown in Figure 8.14. If it has 30 units of capital, for example, its average total cost curve is ATC_{30}. In the long run the firm can examine the average total cost curves associated with varying levels of capital. Four possible short-run average total cost curves for Lifetime Disc are shown in Figure 8.14 for quantities of capital of 20, 30, 40, and 50 units. The relevant curves are labeled ATC_{20}, ATC_{30}, ATC_{40}, and ATC_{50} respectively. The *LRAC* curve is derived from this set of short-run curves by finding the lowest average total cost associated with each level of output. Again, notice that the U-shaped *LRAC* curve is an envelope curve that surrounds the various short-run ATC curves. With the exception of ATC_{40}, in this example, the lowest cost per unit for a particular level of output in the long run is not the minimum point of the relevant short-run curve.

FIGURE 8.14 Relationship Between Short-Run and Long-Run Average Total Costs

The *LRAC* curve is found by taking the lowest average total cost curve at each level of output. Here, average total cost curves for quantities of capital of 20, 30, 40, and 50 units are shown for the Lifetime Disc Co. At a production level of 10,000 CDs per week, Lifetime minimizes its cost per CD by producing with 20 units of capital (point A). At 20,000 CDs per week, an expansion to a plant size associated with 30 units of capital minimizes cost per unit (point B). The lowest cost per unit is achieved with production of 30,000 CDs per week using 40 units of capital (point C). If Lifetime chooses to produce 40,000 CDs per week, it will do so most cheaply with 50 units of capital (point D).



Economies and Diseconomies of Scale

Notice that the long-run average cost curve in Figure 8.14 first slopes downward and then slopes upward. The shape of this curve tells us what is happening to average cost as the firm changes its scale of operations. A firm is said to experience **economies of scale** when long-run average cost declines as the firm expands its output. A firm is said to experience **diseconomies of scale** when long-run average cost increases as the firm expands its output. **Constant returns to scale** occur when long-run average cost stays the same over an output range.

economies of scale

Situation in which the long-run average cost declines as the firm expands its output.

diseconomies of scale

Situation in which the long-run average cost increases as the firm expands its output.

constant returns to scale

Situation in which the long-run average cost stays the same over an output range.

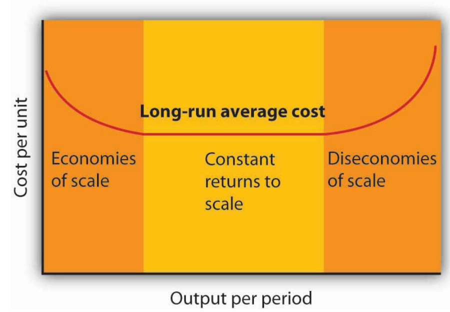
Why would a firm experience economies of scale? One source of economies of scale is gains from specialization. As the scale of a firm's operation expands, it is able to use its factors in more specialized ways, increasing their productivity. Another source of economies of scale lies in the economies that can be gained from mass production methods. As the scale of a firm's operation expands, the company can begin to utilize large-scale machines and production systems that can substantially reduce cost per unit.

Why would a firm experience diseconomies of scale? At first glance, it might seem that the answer lies in the law of diminishing marginal returns, but this is not the case. The law of diminishing marginal returns, after all, tells us how output changes as a single factor is increased, with all other factors of production held constant. In contrast, diseconomies of scale describe a situation of rising average cost even when the firm is free to vary any or all of its factors as it wishes. Diseconomies of scale are generally thought to be caused by management problems. As the scale of a firm's operations expands, it becomes harder and harder for management to coordinate and guide the activities of individual units of the firm. Eventually, the diseconomies of management overwhelm any gains the firm might be achieving by operating with a larger scale of plant, and long-run average costs begin rising. Firms experience constant returns to scale at output levels where there are neither economies nor diseconomies of scale. For the range of output over which the firm experiences constant returns to scale, the long-run average cost curve is horizontal.

Firms are likely to experience all three situations, as shown in Figure 8.15. At very low levels of output, the firm is likely to experience economies of scale as it expands the scale of its operations. There may follow a range of output over which the firm experiences constant returns to scale—empirical studies suggest that the range over which firms experience constant returns to scale is often very large. And certainly there must be some range of output over which diseconomies of scale occur; this phenomenon is one factor that limits the size of firms. A firm operating on the upward-sloping part of its *LRAC* curve is likely to be undercut in the market by smaller firms operating with lower costs per unit of output.

FIGURE 8.15 Economies and Diseconomies of Scale and Long-Run Average Cost

The downward-sloping region of the firm's *LRAC* curve is associated with economies of scale. There may be a horizontal range associated with constant returns to scale. The upward-sloping range of the curve implies diseconomies of scale.



The Size Distribution of Firms

Economies and diseconomies of scale have a powerful effect on the sizes of firms that will operate in any market. Suppose firms in a particular industry experience diseconomies of scale at relatively low levels of output. That industry will be characterized by a large number of fairly small firms. The restaurant market appears to be such an industry. Barbers and beauticians are another example.

If firms in an industry experience economies of scale over a very wide range of output, firms that expand to take advantage of lower cost will force out smaller firms that have higher costs. Such industries are likely to have a few large firms instead of many small ones. In the refrigerator industry, for example, the size of firm necessary to achieve the lowest possible cost per unit is large enough to limit the market to only a few firms. In most cities, economies of scale leave room for only a single newspaper.

One factor that can limit the achievement of economies of scale is the demand facing an individual firm. The scale of output required to achieve the lowest unit costs possible may require sales that exceed the demand facing a firm. A grocery store, for example, could minimize unit costs with a large store and a large volume of sales. But the demand for groceries in a small, isolated community may not be able to sustain such a volume of sales. The firm is thus limited to a small scale of operation even though this might involve higher unit costs.

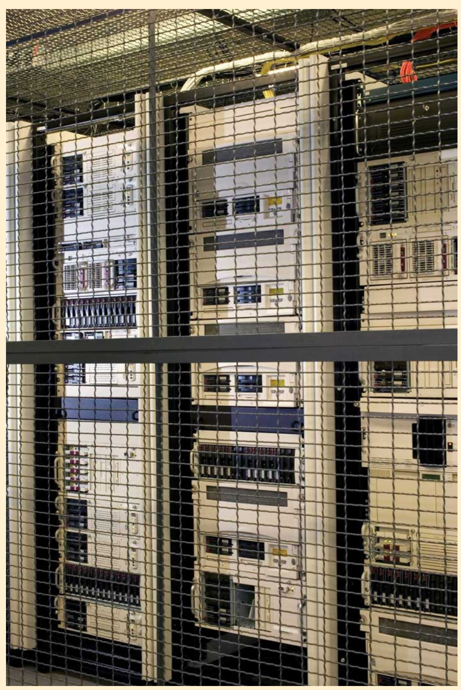
KEY TAKEAWAYS

- A firm chooses its factor mix in the long run on the basis of the marginal decision rule; it seeks to equate the ratio of marginal product to price for all factors of production. By doing so, it minimizes the cost of producing a given level of output.
- The long-run average cost (*LRAC*) curve is derived from the average total cost curves associated with different quantities of the factor that is fixed in the short run. The *LRAC* curve shows the lowest cost per unit at which each quantity can be produced when all factors of production, including capital, are variable.
- A firm may experience economies of scale, constant returns to scale, or diseconomies of scale. Economies of scale imply a downward-sloping long-run average cost (*LRAC*) curve. Constant returns to scale imply a horizontal *LRAC* curve. Diseconomies of scale imply an upward-sloping *LRAC* curve.
- A firm's ability to exploit economies of scale is limited by the extent of market demand for its products.
- The range of output over which firms experience economies of scale, constant return to scale, or diseconomies of scale is an important determinant of how many firms will survive in a particular market.

TRY IT!

1. Suppose Acme Clothing is operating with 20 units of capital and producing 9 units of output at an average total cost of $67, as shown in Figure 8.8. How much labor is it using?
2. Suppose it finds that, with this combination of capital and labor, $MP_K/P_K > MP_L/P_L$. What adjustment will the firm make in the long run? Why does it not make this same adjustment in the short run?

Case in Point: Telecommunications Equipment, Economies of Scale, and Outage Risk



How big should the call switching equipment a major telecommunications company uses be? Having bigger machines results in economies of scale but also raises the risk of larger outages that will affect more customers.

Verizon Laboratories economist Donald E. Smith examined both the economies of scale available from larger equipment and the greater danger of more widespread outages. He concluded that companies should not use the largest machines available because of the outage danger and that they should not use the smallest size because that would mean forgoing the potential gains from economies of scale of larger sizes.

Switching machines, the large computers that handle calls for telecommunications companies, come in four basic "port matrix sizes." These are measured in terms of Digital Cross-Connects (DCS's). The four DCS sizes available are 6,000; 12,000; 24,000; and 36,000 ports. Different machine sizes are made with the same components and thus have essentially the same probability of breaking down. Because larger machines serve more customers, however, a breakdown in a large machine has greater consequences for the company.

The costs of an outage have three elements. The first is lost revenue from calls that would otherwise have been completed. Second, the FCC requires companies to provide a credit of one month of free service after any outage that lasts longer than one minute. Finally, an outage damages a company's reputation and inevitably results in dissatisfied customers—some of whom may switch to other companies.

But, there are advantages to larger machines. A company has a "portfolio" of switching machines. Having larger machines lowers costs in several ways. First, the initial acquisition of the machine generates lower cost per call completed the greater the size of the machine. When the company must make upgrades to the software, having fewer—and larger—machines means fewer upgrades and thus lower costs.

In deciding on matrix size companies should thus compare the cost advantages of a larger matrix with the disadvantages of the higher outage costs associated with those larger matrixes.

Mr. Smith concluded that the economies of scale outweigh the outage risks as a company expands beyond 6,000 ports but that 36,000 ports is "too big" in the sense that the outage costs outweigh the advantage of the economies of scale. The evidence thus suggests that a matrix size in the range of 12,000 to 24,000 ports is optimal.

Source: Donald E. Smith, "How Big Is Too Big? Trading Off the Economies of Scale of Larger Telecommunications Network Elements Against the Risk of Larger Outages," European Journal of Operational Research, 173 (1) (August 2006): 299–312.

ANSWERS TO TRY IT! PROBLEMS

1. To produce 9 jackets, Acme uses 4 units of labor.
2. In the long run, Acme will substitute capital for labor. It cannot make this adjustment in the short run, because its capital is fixed in the short run.

3. REVIEW AND PRACTICE

Summary

In this chapter we have concentrated on the production and cost relationships facing firms in the short run and in the long run.

In the short run, a firm has at least one factor of production that it cannot vary. This fixed factor limits the firm's range of factor choices. As a firm uses more and more of a variable factor (with fixed quantities of other factors of production), it is likely to experience at first increasing, then diminishing, then negative marginal returns. Thus, the short-run total cost curve has a positive value at a zero level of output (the firm's total fixed cost), then slopes upward at a decreasing rate (the range of increasing marginal returns), and then slopes upward at an increasing rate (the range of diminishing marginal returns).

In addition to short-run total product and total cost curves, we derived a firm's marginal product, average product, average total cost, average variable cost, average fixed cost, and marginal cost curves.

If the firm is to maximize profit in the long run, it must select the cost-minimizing combination of factors for its chosen level of output. Thus, the firm must try to use factors of production in accordance with the marginal decision rule. That is, it will use factors so that the ratio of marginal product to factor price is equal for all factors of production.

A firm's long-run average cost (*LRAC*) curve includes a range of economies of scale, over which the curve slopes downward, and a range of diseconomies of scale, over which the curve slopes upward. There may be an intervening range of output over which the firm experiences constant returns to scale; its *LRAC* curve will be horizontal over this range. The size of operations necessary to reach the lowest point on the *LRAC* curve has a great deal to do with determining the relative sizes of firms in an industry.

This chapter has focused on the nature of production processes and the costs associated with them. These ideas will prove useful in understanding the behavior of firms and the decisions they make concerning supply of goods and services.

CONCEPT PROBLEMS

1. Which of the following would be considered long-run choices? Which are short-run choices?
    a. A dentist hires a new part-time dental hygienist.
    b. The local oil refinery plans a complete restructuring of its production processes, including relocating the plant.
    c. A farmer increases the quantity of water applied to his or her fields.
    d. A law partnership signs a 3-year lease for an office complex.
    e. The university hires a new football coach on a 3-year contract.
2. "There are no fixed costs in the long run." Explain.
3. Business is booming at the local McDonald's restaurant. It is contemplating adding a new grill and french-fry machine, but the day supervisor suggests simply hiring more workers. How should the manager decide which alternative to pursue?
4. Suppose that the average age of students in your economics class is 23.7 years. If a new 19-year-old student enrolls in the class, will the average age in the class rise or fall? Explain how this relates to the relationship between average and marginal values.
5. Barry Bond's career home run average in his first 15 years in major league baseball (through 1997) was 33 home runs per season. In 2001, he hit 73 home runs. What happened to his career home run average? What effect did his performance in 2001 have on his career home run average? Explain how this relates to the relationship between average and marginal values.
6. Suppose a firm is operating at the minimum point of its short-run average total cost curve, so that marginal cost equals average total cost. Under what circumstances would it choose to alter the size of its plant? Explain.
7. What happens to the difference between average total cost and average variable cost as a firm's output expands? Explain.
8. How would each of the following affect average total cost, average variable cost, and marginal cost?
    a. An increase in the cost of the lease of the firm's building
    b. A reduction in the price of electricity
    c. A reduction in wages
    d. A change in the salary of the president of the company
9. Consider the following types of firms. For each one, the long-run average cost curve eventually exhibits diseconomies of scale. For which firms would you expect diseconomies of scale to set in at relatively low levels of output? Why?
    a. A copy shop
    b. A hardware store
    c. A dairy
    d. A newspaper
    e. An automobile manufacturer
    f. A restaurant
10. As car manufacturers incorporate more sophisticated computer technology in their vehicles, auto-repair shops require more computerized testing equipment, which is quite expensive, in order to repair newer cars. How is this likely to affect the shape of these firms' long-run average total cost curves? How is it likely to affect the number of auto-repair firms in any market?

NUMERICAL PROBLEMS

1. The table below shows how the number of university classrooms cleaned in an evening varies with the number of janitors:

Janitors per evening	0	1	2	3	4	5	6	7
Classrooms cleaned per evening	0	3	7	12	16	17	17	16

   a. What is the marginal product of the second janitor?
   b. What is the average product of four janitors?
   c. Is the addition of the third janitor associated with increasing, diminishing, or negative marginal returns? Explain.
   d. Is the addition of the fourth janitor associated with increasing, diminishing, or negative marginal returns? Explain.
   e. Is the addition of the seventh janitor associated with increasing, diminishing, or negative marginal returns? Explain.
   f. Draw the total product, average product, and marginal product curves and shade the regions corresponding to increasing marginal returns, decreasing marginal returns, and negative marginal returns.
   g. Calculate the slope of the total product curve as each janitor is added.
   h. Characterize the nature of marginal returns in the region where
      1. The slope of the total product curve is positive and increasing.
      2. The slope of the total product curve is positive and decreasing.
      3. The slope of the total product curve is negative.
2. Suppose a firm is producing 1,000 units of output. Its average fixed costs are $100. Its average variable costs are $50. What is the total cost of producing 1,000 units of output?
3. The director of a nonprofit foundation that sponsors 8-week summer institutes for graduate students analyzed the costs and expected revenues for the next summer institute and recommended that the session be canceled. In her analysis she included a share of the foundation's overhead—the salaries of the director and staff and costs of maintaining the office—to the program. She estimated costs and revenues as follows:

Projected revenues (from tuition and fees)	$300,000
Projected costs	
Overhead	$ 50,000
Room and board for students	$100,000
Costs for faculty and miscellaneous	$175,000
Total costs	$325,000

   What was the error in the director's recommendation?
4. The table below shows the total cost of cleaning classrooms:

Classrooms cleaned per evening	0	3	7	12	16	17
Total cost	$100	$200	$300	$400	$500	$600

   a. What is the average fixed cost of cleaning three classrooms?
   b. What is the average variable cost of cleaning three classrooms?
   c. What is the average fixed cost of cleaning seven classrooms?
   d. What is the average variable cost of cleaning seven classrooms?
   e. What is the marginal cost of cleaning the seventeenth classroom?
   f. What is the average total cost of cleaning twelve classrooms?
5. The average total cost for printing 10,000 copies of an issue of a magazine is $0.45 per copy. For 20,000 copies, the average total cost is $0.35 apiece; for 30,000, the average total cost is $0.30 per copy. The average total cost continues to decline slightly over every level of output that the publishers of the magazine have considered. Sketch the approximate shapes of the average and marginal cost curves. What are some variable costs of publishing magazines? Some fixed costs?
6. The information in the table explains the production of socks. Assume that the price per unit of the variable factor of production (L) is $20 and the price per unit of the fixed factor of production (K) is $5.

Units of Fixed Factor (*K*)	Units of Variable Factor (*L*)	Total Product (*Q*)
10	0	0
10	1	2
10	2	5
10	3	12
10	4	15
10	5	16

a. Add columns to the table and calculate the values for : Marginal Product of Labor (MP_L), Total Variable Cost (*TVC*), Total Fixed Cost (*TFC*), Total Cost (*TC*), Average Variable Cost (*AVC*), Average Fixed Cost (*AFC*), Average Total Cost (*ATC*), and Marginal Cost (*MC*).

b. On two sets of axes, graph the Total Product and Marginal Product curves. Be sure to label curves and axes and remember to plot marginal product using the midpoint convention. Indicate the point on each graph at which diminishing marginal returns appears to begin.

c. Graph Total Variable Cost, Total Fixed Cost, and Total Cost on another set of axes. Indicate the point on the graph at which diminishing marginal returns appears to begin.

d. Graph the Average Fixed Cost, Average Variable Cost, Average Total Cost, and Marginal Cost curves on another set of axes. Indicate the point at which diminishing marginal returns appears to begin.

7. The table below shows the long-run average cost of producing knives:

Knives per hour	1,000	2,000	3,000	4,000	5,000	6,000
Cost per knife	$2	$1.50	$1.00	$1.00	$1.20	$1.30

a. Draw the long-run average cost curve for knives.

b. Shade the regions corresponding to economies of scale, constant returns to scale, and diseconomies of scale.

c. In the region of the long-run average cost curve that corresponds to economies of scale, what is happening to the cost per knife?

d. In the region of the long-run average cost curve that corresponds to constant returns to scale, what is happening to the cost per knife?

e. In the region of the long-run average cost curve that corresponds to diseconomies of scale, what is happening to the cost per knife?

8. Suppose a firm finds that the marginal product of capital is 60 and the marginal product of labor is 20. If the price of capital is $6 and the price of labor is $2.50, how should the firm adjust its mix of capital and labor? What will be the result?

9. A firm minimizes its costs by using inputs such that the marginal product of labor is 10 and the marginal product of capital is 20. The price of capital is $10 per unit. What must the price of labor be?

10. Suppose that the price of labor is $10 per unit and the price of capital is $20 per unit.

a. Assuming the firm is minimizing its cost, if the marginal product of labor is 50, what must the marginal product of capital be?

b. Suppose the price of capital increases to $25 per unit, while the price of labor stays the same. To minimize the cost of producing the same level of output, would the firm become more capital-intensive or labor-intensive? Explain.

ENDNOTES

1. Lucinda Vargas, "Maquiladoras: Impact on Texas Border Cities," in *The Border Economy*, Federal Reserve Bank of Dallas (June 2001): 25–29; William C. Gruben, "Have Mexico's Maquiladoras Bottomed Out?", Southwest Economy, Federal Reserve Bank of Dallas (January/February, 2004), pp. 14–15.

CHAPTER 9

Competitive Markets for Goods and Services

START UP: LIFE ON THE FARM

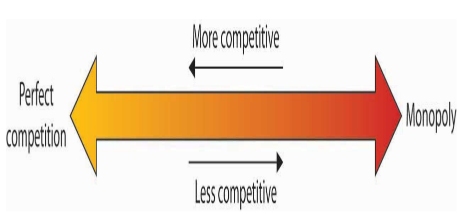
They produce a commodity that is essential to our daily lives, one for which the demand is virtually assured. And yet many—even as farm prices are reaching record highs—seem to live on the margin of failure. Thousands are driven out of business each year. We provide billions of dollars in aid for them, but still we hear of the hardships many of them face. They are our nation's farmers.

What is it about farmers, and farming, that arouses our concern? Much of the answer probably lies in our sense that farming is fundamental to the American way of life. Our country was built, in large part, by independent men and women who made their living from the soil. Many of us perceive their plight as our plight. But part of the answer lies in the fact that farmers do, in fact, face a difficult economic environment. Most of them operate in highly competitive markets, markets that tolerate few mistakes and generally offer small rewards. Finally, perhaps our concern is stirred by our recognition that the farmers' plight is our blessing. The low prices that make life difficult for farmers are the low prices we enjoy as consumers of food.

What keeps the returns to farming as low as they are? What holds many farmers in a situation in which they always seem to be just getting by? In this chapter we shall see that prices just high enough to induce firms to continue to produce are precisely what we would expect to prevail in a competitive market. We will examine a model of how competitive markets work. Not only does this model help to explain the situation facing farmers, but it will also help us to understand the determination of price and output in a wide range of markets. A farm is a firm, and our analysis of such a firm in a competitive market will give us the tools to analyze the choices of all firms operating in competitive markets.

We will put the concepts of marginal cost, average variable cost, and average total cost to work to see how firms in a competitive market respond to market forces. We will see how firms adjust to changes in demand and supply in the short run and in the long run. In all of this, we will be examining how firms use the marginal decision rule.

The competitive model introduced in this chapter lies at one end of a spectrum of market models. At the other end is the monopoly model. It assumes a market in which there is no competition, a market in which only a single firm operates. Two models that fall between the extremes of perfect competition and monopoly are monopolistic competition and oligopoly.



1. PERFECT COMPETITION: A MODEL

LEARNING OBJECTIVES

1. **Explain what economists mean by perfect competition.**
2. **Identify the basic assumptions of the model of perfect competition and explain why they imply price-taking behavior.**

Virtually all firms in a market economy face competition from other firms. In this chapter, we will be working with a model of a highly idealized form of competition called "perfect" by economists.

perfect competition

Model of the market based on the assumption that a large number of firms produce identical goods consumed by a large number of buyers.

Perfect competition is a model of the market based on the assumption that a large number of firms produce identical goods consumed by a large number of buyers. The model of perfect competition also assumes that it is easy for new firms to enter the market and for existing ones to leave. And finally, it assumes that buyers and sellers have complete information about market conditions.

As we examine these assumptions in greater detail, we will see that they allow us to work with the model more easily. No market fully meets the conditions set out in these assumptions. As is always the case with models, our purpose is to understand the way things work, not to describe them. And the model of perfect competition will prove enormously useful in understanding the world of markets.

1.1 Assumptions of the Model

price takers

Individuals or firms who must take the market price as given.

The assumptions of the model of perfect competition, taken together, imply that individual buyers and sellers in a perfectly competitive market accept the market price as given. No one buyer or seller has any influence over that price. Individuals or firms who must take the market price as given are called **price takers**. A consumer or firm that takes the market price as given has no ability to influence that price. A price-taking firm or consumer is like an individual who is buying or selling stocks. He or she looks up the market price and buys or sells at that price. The price is determined by demand and supply in the market—not by individual buyers or sellers. In a perfectly competitive market, each firm and each consumer is a price taker. A price-taking consumer assumes that he or she can purchase any quantity at the market price—without affecting that price. Similarly, a price-taking firm assumes it can sell whatever quantity it wishes at the market price without affecting the price.

You are a price taker when you go into a store. You observe the prices listed and make a choice to buy or not. Your choice will not affect that price. Should you sell a textbook back to your campus bookstore at the end of a course, you are a price-taking seller. You are confronted by a market price and you decide whether to sell or not. Your decision will not affect that price.

To see how the assumptions of the model of perfect competition imply price-taking behavior, let us examine each of them in turn.

Identical Goods

In a perfectly competitive market for a good or service, one unit of the good or service cannot be differentiated from any other on any basis. A bushel of, say, hard winter wheat is an example. A bushel produced by one farmer is identical to that produced by another. There are no brand preferences or consumer loyalties.

The assumption that goods are identical is necessary if firms are to be price takers. If one farmer's wheat were perceived as having special properties that distinguished it from other wheat, then that farmer would have some power over its price. By assuming that all goods and services produced by firms in a perfectly competitive market are identical, we establish a necessary condition for price-taking behavior. Economists sometimes say that the goods or services in a perfectly competitive market are *homogeneous*, meaning that they are all alike. There are no brand differences in a perfectly competitive market.

A Large Number of Buyers and Sellers

How many buyers and sellers are in our market? The answer rests on our presumption of price-taking behavior. There are so many buyers and sellers that none of them has any influence on the market price regardless of how much any of them purchases or sells. A firm in a perfectly competitive market can react to prices, but cannot affect the prices it pays for the factors of production or the prices it receives for its output.

Ease of Entry and Exit

The assumption that it is easy for other firms to enter a perfectly competitive market implies an even greater degree of competition. Firms in a market must deal not only with the large number of competing firms but also with the possibility that still more firms might enter the market.

Later in this chapter, we will see how ease of entry is related to the sustainability of economic profits. If entry is easy, then the promise of high economic profits will quickly attract new firms. If entry is difficult, it won't.

The model of perfect competition assumes easy exit as well as easy entry. The assumption of easy exit strengthens the assumption of easy entry. Suppose a firm is considering entering a particular market. Entry may be easy, but suppose that getting out is difficult. For example, suppliers of factors of production to firms in the industry might be happy to accommodate new firms but might require that they sign long-term contracts. Such contracts could make leaving the market difficult and costly. If that were the case, a firm might be hesitant to enter in the first place. Easy exit helps make entry easier.

Complete Information

We assume that all sellers have complete information about prices, technology, and all other knowledge relevant to the operation of the market. No one seller has any information about production methods that is not available to all other sellers. If one seller had an advantage over other sellers, perhaps special information about a lower-cost production method, then that seller could exert some control over market price—the seller would no longer be a price taker.

We assume also that buyers know the prices offered by every seller. If buyers did not know about prices offered by different firms in the market, then a firm might be able to sell a good or service for a price other than the market price and thus could avoid being a price taker.

The availability of information that is assumed in the model of perfect competition implies that information can be obtained at low cost. If consumers and firms can obtain information at low cost, they are likely to do so. Information about the marketplace may come over the internet, over the airways in a television commercial, or over a cup of coffee with a friend. Whatever its source, we assume that its low cost ensures that consumers and firms have enough of it so that everyone buys or sells goods and services at market prices determined by the intersection of demand and supply curves.

The assumptions of the perfectly competitive model ensure that each buyer or seller is a price taker. The market, not individual consumers or firms, determines price in the model of perfect competition. No individual has enough power in a perfectly competitive market to have any impact on that price.

1.2 Perfect Competition and the Real World

The assumptions of identical products, a large number of buyers, easy entry and exit, and perfect information are strong assumptions. The notion that firms must sit back and let the market determine price seems to fly in the face of what we know about most real firms, which is that firms customarily do set prices. Yet this is the basis for the model of demand and supply, the power of which you have already seen.

When we use the model of demand and supply, we assume that market forces determine prices. In this model, buyers and sellers respond to the market price. They are price takers. The assumptions of the model of perfect competition underlie the assumption of price-taking behavior. Thus we are using the model of perfect competition whenever we apply the model of demand and supply.

We can understand most markets by applying the model of demand and supply. Even though those markets do not fulfill all the assumptions of the model of perfect competition, the model allows us to understand some key features of these markets.

Changes within your lifetime have made many markets more competitive. Falling costs of transportation, together with dramatic advances in telecommunications, have opened the possibility of entering markets to firms all over the world. A company in South Korea can compete in the market for steel in the United States. A furniture maker in New Mexico can compete in the market for furniture in Japan. A firm can enter the world market simply by creating a web page to advertise its products and to take orders.

In the remaining sections of this chapter, we will learn more about the response of firms to market prices. We will see how firms respond, in the short run and in the long run, to changes in demand and to changes in production costs. In short, we will be examining the forces that constitute the supply side of the model of demand and supply.

We will also see how competitive markets work to serve consumer interests and how competition acts to push economic profits down, sometimes eliminating them entirely. When we have finished we will have a better understanding of the market conditions facing farmers and of the conditions that prevail in any competitive industry.

KEY TAKEAWAYS

- The central characteristic of the model of perfect competition is the fact that price is determined by the interaction of demand and supply; buyers and sellers are price takers.
- The model assumes: a large number of firms producing identical (homogeneous) goods or services, a large number of buyers and sellers, easy entry and exit in the industry, and complete information about prices in the market.
- The model of perfect competition underlies the model of demand and supply.

TRY IT!

Which of the following goods and services are likely produced in a perfectly competitive industry? Relate your answer to the assumptions of the model of perfect competition.

1. International express mail service
2. Corn
3. Athletic shoes

Case in Point: Entering and Exiting the Burkha Industry

Muhammed Ibrahim Islamadin was driving a cab in Kabul, Afghanistan, when the Taliban took over the country. He foresaw the repression that would follow and sensed an opportunity.

He sold his taxicab and set up a shop for sewing and selling burkhas, the garments required of all women under the Taliban's rule. Mr. Islamadin had an easy task selling, as women caught outdoors with exposed skin were routinely beaten by the Taliban's religious police. He told *The Wall Street Journal*, "This was very bad for them, but it was good for me."

Of course, Mr. Islamadin was not the only producer to get into the industry. Other Afghani merchants, as well as merchants from Pakistan and China, also jumped at the opportunity.

The entry of new firms exemplifies an important characteristic of perfect competition. Whenever there is an opportunity to earn economic profits—even an unexpected opportunity—new firms will enter, provided that entry is easy.

The model of perfect competition also assumes that exit will be easy if and when a firm experiences economic losses. When the Taliban rulers were ousted by the United States and its allies in 2001, Mr. Islamadin expected that the demand for burkhas would begin to fall. It did. The sales fell 50% almost immediately. Prices fell as well, generally by about 20%.

It was simple for Mr. Islamadin to leave the industry. He gave his remaining stock of burkhas to a brother who was producing them in the countryside where women continued to wear them. As for Mr. Islamadin, he has made plans to go into the glassware business. He expects the demand for glass teacups to be strong whatever happens in Afghanistan's critical future.

Source: Andrew Higgins, "With Islamic Dress, Out Goes the Guy Who Sold Burkhas," The Wall Street Journal, December 19, 2001, p. A1.

ANSWERS TO TRY IT! PROBLEMS

1. Not perfectly competitive–There are few sellers in this market (Fedex, UPS, and the United States Postal Services are the main ones in the United States) probably because of the difficulty of entry and exit. To provide these services requires many outlets and a large transportation fleet, for example.
2. Perfectly competitive—There are many firms producing a largely homogeneous product and there is good information about prices. Entry and exit is also fairly easy as firms can switch among a variety of crops.
3. Not perfectly competitive—The main reason is that goods are not identical.

2. OUTPUT DETERMINATION IN THE SHORT RUN

LEARNING OBJECTIVES

1. **Show graphically how an individual firm in a perfectly competitive market can use total revenue and total cost curves or marginal revenue and marginal cost curves to determine the level of output that will maximize its economic profit.**
2. **Explain when a firm will shut down in the short run and when it will operate even if it is incurring economic losses.**
3. **Derive the firm's supply curve from the firm's marginal cost curve and the industry supply curve from the supply curves of individual firms.**

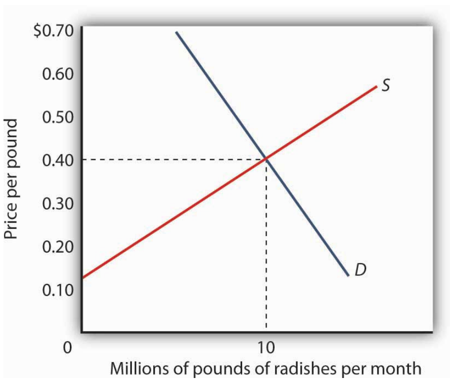
Our goal in this section is to see how a firm in a perfectly competitive market determines its output level in the short run—a planning period in which at least one factor of production is fixed in quantity. We shall see that the firm can maximize economic profit by applying the marginal decision rule and increasing output up to the point at which the marginal benefit of an additional unit of output is just equal to the marginal cost. This fact has an important implication: over a wide range of output, the firm's marginal cost curve is its supply curve.

2.1 Price and Revenue

Each firm in a perfectly competitive market is a price taker; the equilibrium price and industry output are determined by demand and supply. Figure 9.3 shows how demand and supply in the market for radishes, which we shall assume are produced under conditions of perfect competition, determine total output and price. The equilibrium price is $0.40 per pound; the equilibrium quantity is 10 million pounds per month.

FIGURE 9.3 **The Market for Radishes**

Price and output in a competitive market are determined by demand and supply. In the market for radishes, the equilibrium price is \$0.40 per pound; 10 million pounds per month are produced and purchased at this price.



Because it is a price taker, each firm in the radish industry assumes it can sell all the radishes it wants at a price of \$0.40 per pound. No matter how many or how few radishes it produces, the firm expects to sell them all at the market price.

The assumption that the firm expects to sell all the radishes it wants at the market price is crucial. If a firm did not expect to sell all of its radishes at the market price—if it had to lower the price to sell some quantities—the firm would not be a price taker. And price-taking behavior is central to the model of perfect competition.

Radish growers—and perfectly competitive firms in general—have no reason to charge a price lower than the market price. Because buyers have complete information and because we assume each firm's product is identical to that of its rivals, firms are unable to charge a price higher than the market price. For perfectly competitive firms, the price is very much like the weather: they may complain about it, but in perfect competition there is nothing any of them can do about it.

Total Revenue

While a firm in a perfectly competitive market has no influence over its price, it does determine the output it will produce. In selecting the quantity of that output, one important consideration is the revenue the firm will gain by producing it.

total revenue

A firm's output multiplied by the price at which it sells that output.

A firm's **total revenue** is found by multiplying its output by the price at which it sells that output. For a perfectly competitive firm, total revenue (TR) is the market price (P) times the quantity the firm produces (Q), or

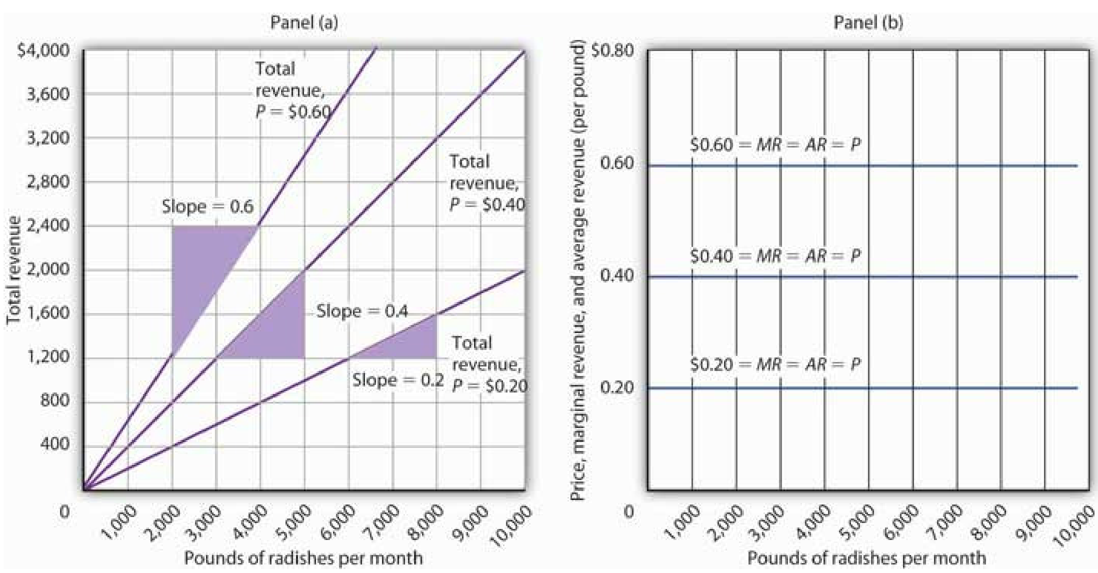
EQUATION 9.1

$$TR = P \times Q$$

The relationship between market price and the firm's total revenue curve is a crucial one. Panel (a) of Figure 9.4 shows total revenue curves for a radish grower at three possible market prices: \$0.20, \$0.40, and \$0.60 per pound. Each total revenue curve is a linear, upward-sloping curve. At any price, the greater the quantity a perfectly competitive firm sells, the greater its total revenue. Notice that the greater the price, the steeper the total revenue curve is.

FIGURE 9.4 **Total Revenue, Marginal Revenue, and Average Revenue**

Panel (a) shows different total revenue curves for three possible market prices in perfect competition. A total revenue curve is a straight line coming out of the origin. The slope of a total revenue curve is *MR;* it equals the market price (*P*) and *AR* in perfect competition. Marginal revenue and average revenue are thus a single horizontal line at the market price, as shown in Panel (b). There is a different marginal revenue curve for each price.



Price, Marginal Revenue, and Average Revenue

The slope of a total revenue curve is particularly important. It equals the change in the vertical axis (total revenue) divided by the change in the horizontal axis (quantity) between any two points. The slope measures the rate at which total revenue increases as output increases. We can think of it as the increase in total revenue associated with a 1-unit increase in output. The increase in total revenue from a 1-unit increase in quantity is **marginal revenue**. Thus marginal revenue (*MR*) equals the slope of the total revenue curve.

marginal revenue

The increase in total revenue from a one-unit increase in quantity.

How much additional revenue does a radish producer gain from selling one more pound of radishes? The answer, of course, is the market price for 1 pound. Marginal revenue equals the market price. Because the market price is not affected by the output choice of a single firm, the marginal revenue the firm gains by producing one more unit is always the market price. The marginal revenue curve shows the relationship between marginal revenue and the quantity a firm produces. For a perfectly competitive firm, the marginal revenue curve is a horizontal line at the market price. If the market price of a pound of radishes is \$0.40, then the marginal revenue is \$0.40. Marginal revenue curves for prices of \$0.20, \$0.40, and \$0.60 are given in Panel (b) of Figure 9.4. In perfect competition, a firm's marginal revenue curve is a horizontal line at the market price.

Price also equals **average revenue**, which is total revenue divided by quantity. Equation 9.1 gives total revenue, *TR*. To obtain average revenue (*AR*), we divide total revenue by quantity, *Q*. Because total revenue equals price (*P*) times quantity (*Q*), dividing by quantity leaves us with price.

average revenue

Total revenue divided by quantity.

EQUATION 9.2

$$AR = \frac{TR}{Q} = \frac{P \times Q}{Q} = P$$

The marginal revenue curve is a horizontal line at the market price, and average revenue equals the market price. The average and marginal revenue curves are given by the same horizontal line. This is consistent with what we have learned about the relationship between marginal and average values. When the marginal value exceeds the average value, the average value will be rising. When the marginal value is less than the average value, the average value will be falling. What happens when the average and marginal values do not change, as in the horizontal curves of Panel (b) of Figure 9.4? The marginal value must equal the average value; the two curves coincide.

Marginal Revenue, Price, and Demand for the Perfectly Competitive Firm

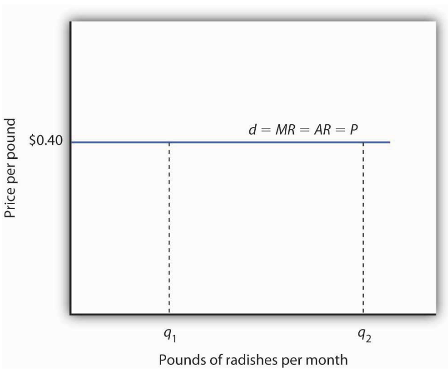
We have seen that a perfectly competitive firm's marginal revenue curve is simply a horizontal line at the market price and that this same line is also the firm's average revenue curve. For the perfectly competitive firm, $MR = P = AR$. The marginal revenue curve has another meaning as well. It is the demand curve facing a perfectly competitive firm.

Consider the case of a single radish producer, Tony Gortari. We assume that the radish market is perfectly competitive; Mr. Gortari runs a perfectly competitive firm. Suppose the market price of radishes is \$0.40 per pound. How many pounds of radishes can Mr. Gortari sell at this price? The answer comes from our assumption that he is a price taker: He can sell *any* quantity he wishes at this price. How many pounds of radishes will he sell if he charges a price that exceeds the market price? None. His radishes are identical to those of every other firm in the market, and everyone in the market has complete information. That means the demand curve facing Mr. Gortari is a horizontal line at the market price as illustrated in Figure 9.5. Notice that the curve is labeled *d* to distinguish it from the market demand curve, *D*, in Figure 9.3. The horizontal line in Figure 9.5 is also Mr. Gortari's marginal revenue curve, *MR*, and his average revenue curve, *AR*. It is also the market price, *P*.

Of course, Mr. Gortari could charge a price below the market price, but why would he? We assume he can sell all the radishes he wants at the market price; there would be no reason to charge a lower price. Mr. Gortari faces a demand curve that is a horizontal line at the market price. In our subsequent analysis, we shall refer to the horizontal line at the market price simply as marginal revenue. We should remember, however, that this same line gives us the market price, average revenue, and the demand curve facing the firm.

FIGURE 9.5 Price, Marginal Revenue, and Demand

A perfectly competitive firm faces a horizontal demand curve at the market price. Here, radish grower Tony Gortari faces demand curve *d* at the market price of $0.40 per pound. He could sell q_1 or q_2—or any other quantity—at a price of $0.40 per pound.



More generally, we can say that *any* perfectly competitive firm faces a horizontal demand curve at the market price. We saw an example of a horizontal demand curve in the chapter on elasticity. Such a curve is perfectly elastic, meaning that any quantity is demanded at a given price.

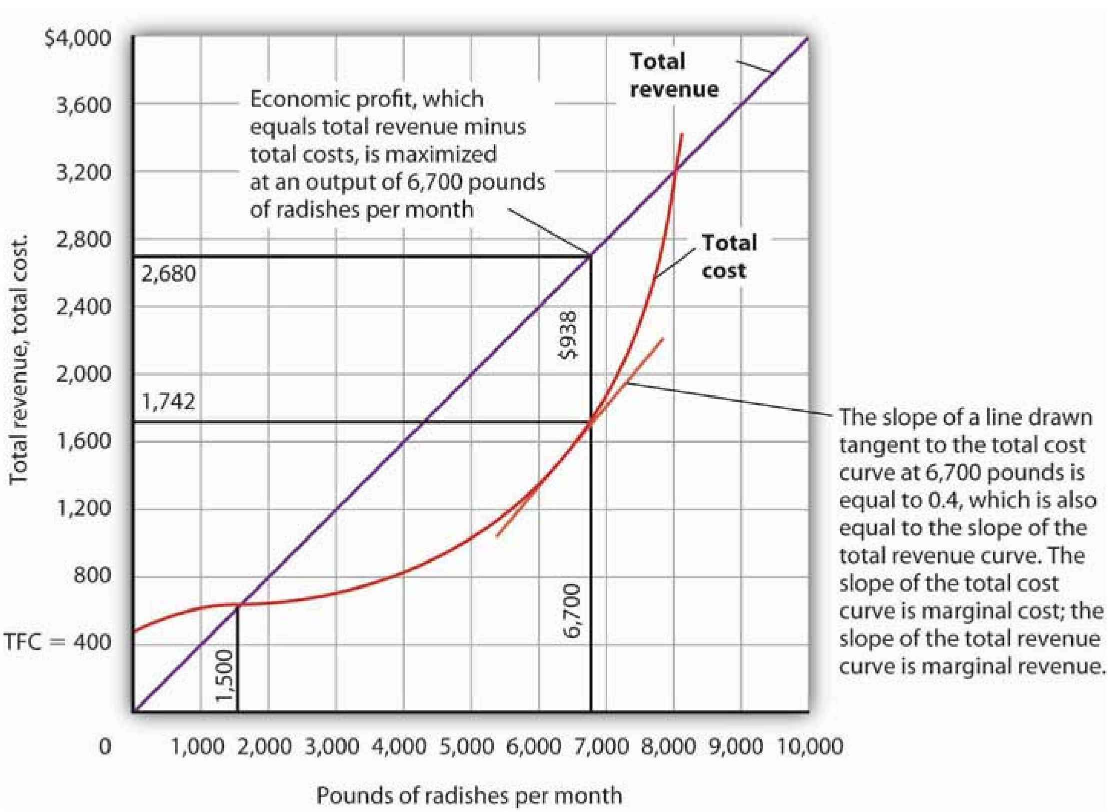
2.2 Economic Profit in the Short Run

A firm's economic profit is the difference between total revenue and total cost. Recall that total cost is the opportunity cost of producing a certain good or service. When we speak of economic profit we are speaking of a firm's total revenue less the total opportunity cost of its operations.

As we learned, a firm's total cost curve in the short run intersects the vertical axis at some positive value equal to the firm's total fixed costs. Total cost then rises at a decreasing rate over the range of increasing marginal returns to the firm's variable factors. It rises at an increasing rate over the range of diminishing marginal returns. Figure 9.6 shows the total cost curve for Mr. Gortari, as well as the total revenue curve for a price of $0.40 per pound. Suppose that his total fixed cost is $400 per month. For any given level of output, Mr. Gortari's economic profit is the vertical distance between the total revenue curve and the total cost curve at that level.

FIGURE 9.6 Total Revenue, Total Cost, and Economic Profit

Economic profit is the vertical distance between the total revenue and total cost curves (revenue minus costs). Here, the maximum profit attainable by Tony Gortari for his radish production is $938 per month at an output of 6,700 pounds.



Let us examine the total revenue and total cost curves in Figure 9.6 more carefully. At zero units of output, Mr. Gortari's total cost is $400 (his total fixed cost); total revenue is zero. Total cost continues to exceed total revenue up to an output of 1,500 pounds per month, at which point the two curves intersect. At this point, economic profit equals zero. As Mr. Gortari expands output above 1,500 pounds per month, total revenue becomes greater than total cost. We see that at a quantity of 1,500 pounds per month, the total revenue curve is steeper than the total cost curve. Because revenues are rising faster than costs, profits rise with increased output. As long as the total revenue curve is steeper than the total cost curve, profit increases as the firm increases its output.

The total revenue curve's slope does not change as the firm increases its output. But the total cost curve becomes steeper and steeper as diminishing marginal returns set in. Eventually, the total cost and total revenue curves will have the same slope. That happens in Figure 9.6 at an output of 6,700 pounds of radishes per month. Notice that a line drawn tangent to the total cost curve at that quantity has the same slope as the total revenue curve.

As output increases beyond 6,700 pounds, the total cost curve continues to become steeper. It becomes steeper than the total revenue curve, and profits fall as costs rise faster than revenues. At an output slightly above 8,000 pounds per month, the total revenue and cost curves intersect again, and economic profit equals zero. Mr. Gortari achieves the greatest profit possible by producing 6,700 pounds of radishes per month, the quantity at which the total cost and total revenue curves have the same slope. More generally, we can conclude that a perfectly competitive firm maximizes economic profit at the output level at which the total revenue curve and the total cost curve have the same slope.

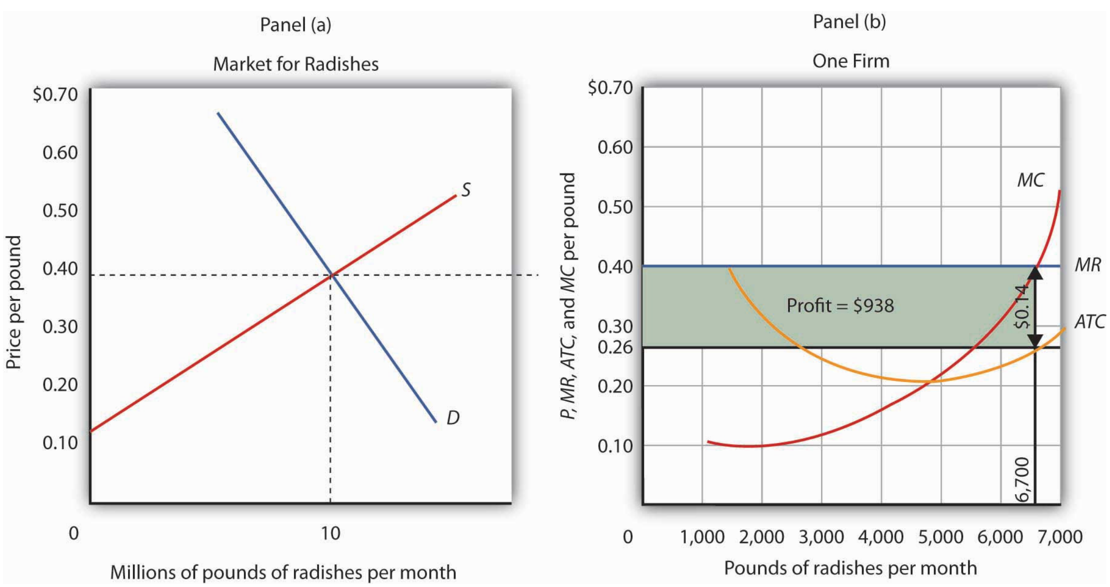
2.3 Applying the Marginal Decision Rule

The slope of the total revenue curve is marginal revenue; the slope of the total cost curve is marginal cost. Economic profit, the difference between total revenue and total cost, is maximized where marginal revenue equals marginal cost. This is consistent with the marginal decision rule, which holds that a profit-maximizing firm should increase output until the marginal benefit of an additional unit equals the marginal cost. The marginal benefit of selling an additional unit is measured as marginal revenue. Finding the output at which marginal revenue equals marginal cost is thus an application of our marginal decision rule.

Figure 9.7 shows how a firm can use the marginal decision rule to determine its profit-maximizing output. Panel (a) shows the market for radishes; the market demand curve (*D*), and supply curve (*S*) that we had in Figure 9.3; the market price is \$0.40 per pound. In Panel (b), the *MR* curve is given by a horizontal line at the market price. The firm's marginal cost curve (*MC*) intersects the marginal revenue curve at the point where profit is maximized. Mr. Gortari maximizes profits by producing 6,700 pounds of radishes per month. That is, of course, the result we obtained in Figure 9.6, where we saw that the firm's total revenue and total cost curves differ by the greatest amount at the point at which the slopes of the curves, which equal marginal revenue and marginal cost, respectively, are equal.

FIGURE 9.7 Applying the Marginal Decision Rule

The market price is determined by the intersection of demand and supply. As always, the firm maximizes profit by applying the marginal decision rule. It takes the market price, \$0.40 per pound, as given and selects an output at which *MR* equals *MC*. Economic profit per unit is the difference between *ATC* and price (here, \$0.14 per pound); economic profit is profit per unit times the quantity produced (\$0.14×6,700=\$938).



economic profit per unit

The difference between price and average total cost.

We can use the graph in Figure 9.7 to compute Mr. Gortari's economic profit. **Economic profit per unit** is the difference between price and average total cost. At the profit-maximizing output of 6,700 pounds of radishes per month, average total cost (*ATC*) is $0.26 per pound, as shown in Panel (b). Price is $0.40 per pound, so economic profit per unit is $0.14. Economic profit is found by multiplying economic profit per unit by the number of units produced; the firm's economic profit is thus $938 ($0.14 × 6,700). It is shown graphically by the area of the shaded rectangle in Panel (b); this area equals the vertical distance between marginal revenue (*MR*) and average total cost (*ATC*) at an output of 6,700 pounds of radishes times the number of pounds of radishes produced, 6,700, in Figure 9.7.

Heads Up!

Look carefully at the rectangle that shows economic profit in Panel (b) of Figure 9.7. It is found by taking the profit-maximizing quantity, 6,700 pounds, then reading up to the *ATC* curve and the firm's demand curve at the market price. Economic profit per unit equals price minus average total cost (*P* − *ATC*).

The firm's economic profit equals economic profit per unit times the quantity produced. It is found by extending horizontal lines from the *ATC* and *MR curve* to the vertical axis and taking the area of the rectangle formed.

There is no reason for the profit-maximizing quantity to correspond to the lowest point on the *ATC* curve; it does not in this case. Students sometimes make the mistake of calculating economic profit as the difference between the price and the lowest point on the *ATC* curve. That gives us the maximum economic profit per unit, but we assume that firms maximize economic profit, not economic profit per unit. The firm's economic profit equals economic profit per unit times quantity. The quantity that maximizes economic profit is determined by the intersection of *ATC* and *MR*.

2.4 Economic Losses in the Short Run

economic loss

The amount by which a firm's total cost exceeds its total revenue.

In the short run, a firm has one or more inputs whose quantities are fixed. That means that in the short run the firm cannot leave its industry. Even if it cannot cover *all* of its costs, including both its variable and fixed costs, going entirely out of business is not an option in the short run. The firm may close its doors, but it must continue to pay its fixed costs. It is forced to accept an **economic loss**, the amount by which its total cost exceeds its total revenue.

Suppose, for example, that a manufacturer has signed a 1-year lease on some equipment. It must make payments for this equipment during the term of its lease, whether it produces anything or not. During the period of the lease, the payments represent a fixed cost for the firm.

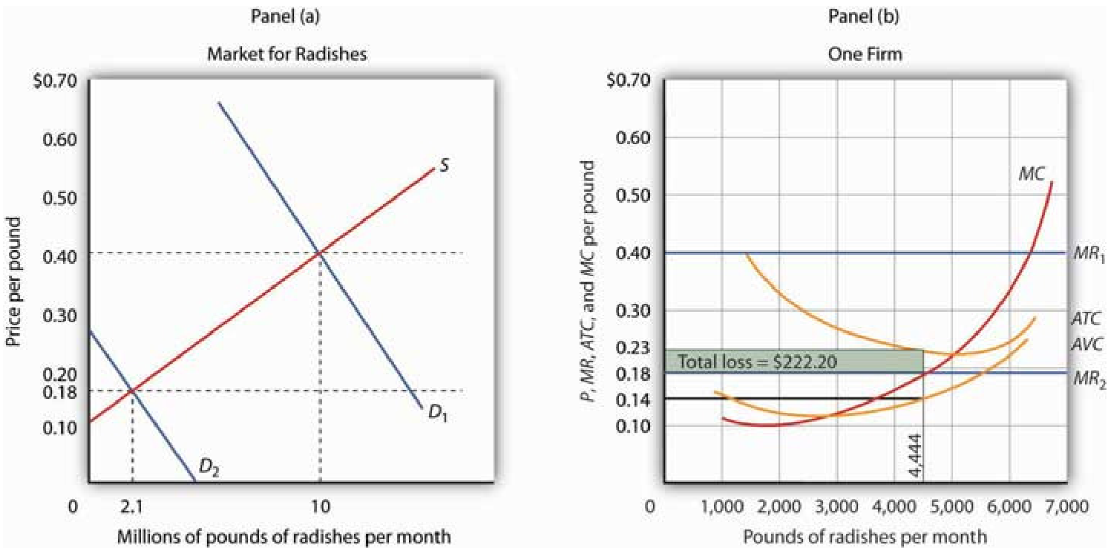
A firm that is experiencing economic losses—whose economic profits have become negative—in the short run may either continue to produce or shut down its operations, reducing its output to zero. It will choose the option that minimizes its losses. The crucial test of whether to operate or shut down lies in the relationship between price and average variable cost.

Producing to Minimize Economic Loss

Suppose the demand for radishes falls to D_2, as shown in Panel (a) of Figure 9.8. The market price for radishes plunges to $0.18 per pound, which is below average total cost. Consequently Mr. Gortari experiences negative economic profits—a loss. Although the new market price falls short of average total cost, it still exceeds average variable cost, shown in Panel (b) as *AVC*. Therefore, Mr. Gortari should continue to produce an output at which marginal cost equals marginal revenue. These curves (labeled *MC* and MR_2) intersect in Panel (b) at an output of 4,444 pounds of radishes per month.

FIGURE 9.8 Suffering Economic Losses in the Short Run

Tony Gortari experiences a loss when price drops below *ATC*, as it does in Panel (b) as a result of a reduction in demand. If price is above *AVC*, however, he can minimize his losses by producing where *MC* equals MR_2. Here, that occurs at an output of 4,444 pounds of radishes per month. The price is $0.18 per pound, and average total cost is $0.23 per pound. He loses $0.05 per pound, or $222.20 per month.



When producing 4,444 pounds of radishes per month, Mr. Gortari faces an average total cost of $0.23 per pound. At a price of $0.18 per pound, he loses a nickel on each pound produced. Total economic losses at an output of 4,444 pounds per month are thus $222.20 per month (=4,444×$0.05).

No producer likes a loss (that is, negative economic profit), but the loss solution shown in Figure 9.8 is the best Mr. Gortari can attain. Any level of production other than the one at which marginal cost equals marginal revenue would produce even greater losses.

Suppose Mr. Gortari were to shut down and produce no radishes. Ceasing production would reduce variable costs to zero, but he would still face fixed costs of $400 per month (recall that $400 was the vertical intercept of the total cost curve in Figure 9.6). By shutting down, Mr. Gortari would lose $400 per month. By continuing to produce, he loses only $222.20.

Mr. Gortari is better off producing where marginal cost equals marginal revenue because at that output price exceeds average variable cost. Average variable cost is $0.14 per pound, so by continuing to produce he covers his variable costs, with $0.04 per pound left over to apply to fixed costs. Whenever price is greater than average variable cost, the firm maximizes economic profit (or minimizes economic loss) by producing the output level at which marginal revenue and marginal cost curves intersect.

Shutting Down to Minimize Economic Loss

Suppose price drops below a firm's average variable cost. Now the best strategy for the firm is to shut down, reducing its output to zero. The minimum level of average variable cost, which occurs at the intersection of the marginal cost curve and the average variable cost curve, is called the **shutdown point**. Any price below the minimum value of average variable cost will cause the firm to shut down. If the firm were to continue producing, not only would it lose its fixed costs, but it would also face an additional loss by not covering its variable costs.

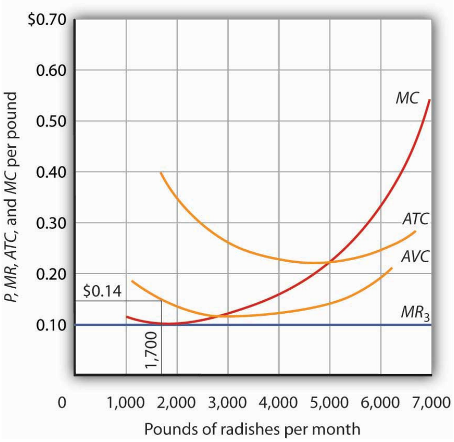
Figure 9.9 shows a case where the price of radishes drops to $0.10 per pound. Price is less than average variable cost, so Mr. Gortari not only would lose his fixed cost but would also incur additional losses by producing. Suppose, for example, he decided to operate where marginal cost equals marginal revenue, producing 1,700 pounds of radishes per month. Average variable cost equals $0.14 per pound, so he would lose $0.04 on each pound he produces ($68) plus his fixed cost of $400 per month. He would lose $468 per month. If he shut down, he would lose only his fixed cost. Because the price of $0.10 falls below his average variable cost, his best course would be to shut down.

shutdown point

The minimum level of average variable cost, which occurs at the intersection of the marginal cost curve and the average variable cost curve.

FIGURE 9.9 Shutting Down

The market price of radishes drops to $0.10 per pound, so MR_3 is below Mr. Gortari's *AVC*. Thus he would suffer a greater loss by continuing to operate than by shutting down. Whenever price falls below average variable cost, the firm will shut down, reducing its production to zero.



Shutting down is not the same thing as going out of business. A firm shuts down by closing its doors; it can reopen them whenever it expects to cover its variable costs. We can even think of a firm's decision to close at the end of the day as a kind of shutdown point; the firm makes this choice because it does not anticipate that it will be able to cover its variable cost overnight. It expects to cover those costs the next morning when it reopens its doors.

2.5 Marginal Cost and Supply

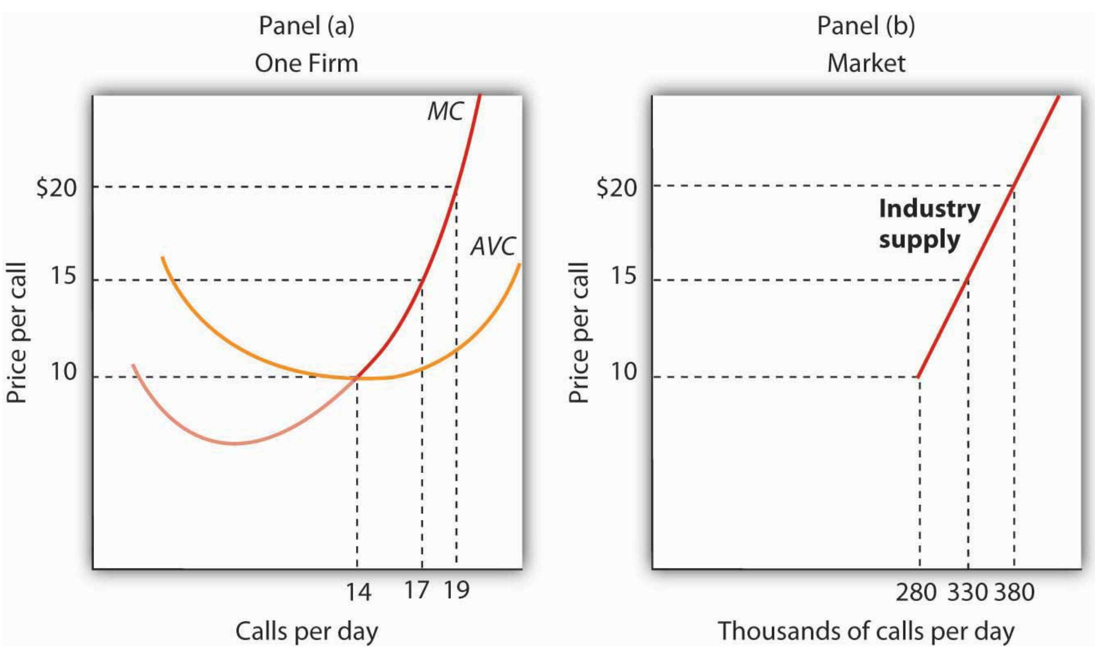
In the model of perfect competition, we assume that a firm determines its output by finding the point where the marginal revenue and marginal cost curves intersect. Provided that price exceeds average variable cost, the firm produces the quantity determined by the intersection of the two curves.

A supply curve tells us the quantity that will be produced at each price, and that is what the firm's marginal cost curve tells us. The firm's supply curve in the short run is its marginal cost curve for prices above the average variable cost. At prices below average variable cost, the firm's output drops to zero.

Panel (a) of Figure 9.10 shows the average variable cost and marginal cost curves for a hypothetical astrologer, Madame LaFarge, who is in the business of providing astrological consultations over the telephone. We shall assume that this industry is perfectly competitive. At any price below $10 per call, Madame LaFarge would shut down. If the price is $10 or greater, however, she produces an output at which price equals marginal cost. The marginal cost curve is thus her supply curve at all prices greater than $10.

FIGURE 9.10 Marginal Cost and Supply

The supply curve for a firm is that portion of its *MC* curve that lies above the *AVC* curve, shown in Panel (a). To obtain the short-run supply curve for the industry, we add the outputs of each firm at each price. The industry supply curve is given in Panel (b).



Now suppose that the astrological forecast industry consists of Madame LaFarge and thousands of other firms similar to hers. The market supply curve is found by adding the outputs of each firm at each price, as shown in Panel (b) of Figure 9.10. At a price of $10 per call, for example, Madame LaFarge supplies 14 calls per day. Adding the quantities supplied by all the other firms in the market, suppose we get a quantity supplied of 280,000. Notice that the market supply curve we have drawn is linear; throughout the book we have made the assumption that market demand and supply curves are linear in order to simplify our analysis.

Looking at Figure 9.10, we see that profit-maximizing choices by firms in a perfectly competitive market will generate a market supply curve that reflects marginal cost. Provided there are no external benefits or costs in producing a good or service, a perfectly competitive market satisfies the efficiency condition.

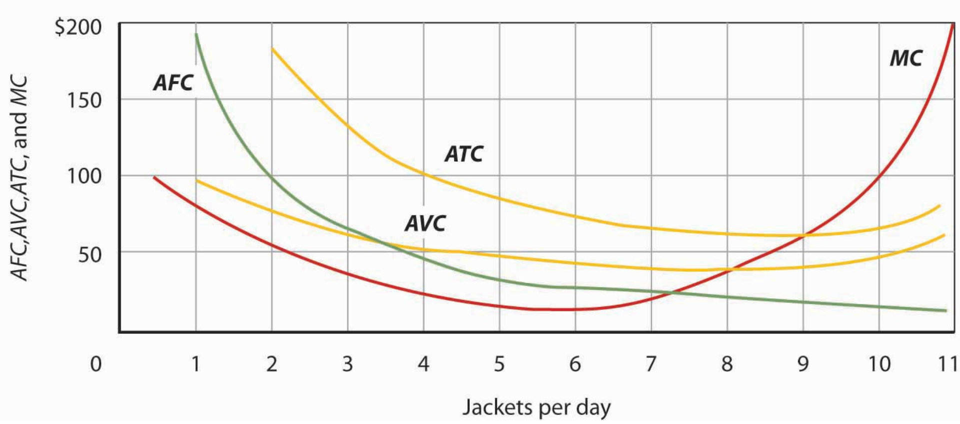
KEY TAKEAWAYS

- Price in a perfectly competitive industry is determined by the interaction of demand and supply.
- In a perfectly competitive industry, a firm's total revenue curve is a straight, upward-sloping line whose slope is the market price. Economic profit is maximized at the output level at which the slopes of the total revenue and total cost curves are equal, provided that the firm is covering its variable cost.
- To use the marginal decision rule in profit maximization, the firm produces the output at which marginal cost equals marginal revenue. Economic profit per unit is price minus average total cost; total economic profit equals economic profit per unit times quantity.
- If price falls below average total cost, but remains above average variable cost, the firm will continue to operate in the short run, producing the quantity where $MR = MC$ doing so minimizes its losses.
- If price falls below average variable cost, the firm will shut down in the short run, reducing output to zero. The lowest point on the average variable cost curve is called the shutdown point.
- The firm's supply curve in the short run is its marginal cost curve for prices greater than the minimum average variable cost.

TRY IT!

Assume that Acme Clothing, the firm introduced in the chapter on production and cost, produces jackets in a perfectly competitive market. Suppose the demand and supply curves for jackets intersect at a price of $81. Now, using the marginal cost and average total cost curves for Acme shown here:

Quantity/day	0	1	2	3	4	5	6	7	8	9	10	11
Total cost	$200	$300	$363	$400	$433	$458	$480	$500	$538	$600	$700	$900
AFC		$200	$100	$67	$50	$40	$33	$29	$25	$22	$20	$18
AVC		$100	$82	$67	$58	$52	$47	$43	$42	$44	$50	$64
ATC		$300	$182	$133	$108	$92	$80	$71	$67	$67	$70	$82
MC		$100	$63	$37	$33	$25	$22	$20	$38	$62	$100	$200



Estimate Acme's profit-maximizing output per day (assume the firm selects a whole number). What are Acme's economic profits per day?

Case in Point: Not Out of Business 'Til They Fall from the Sky

The 66 satellites were poised to start falling from the sky. The hope was that the pieces would burn to bits on their way down through the atmosphere, but there was the chance that a building or a person would take a direct hit.

The satellites were the primary communication devices of Iridium's satellite phone system. Begun in 1998 as the first truly global satellite system for mobile phones—providing communications across deserts, in the middle of oceans, and at the poles—Iridium expected five million subscribers to pay $7 a minute to talk on $3,000 handsets. In the climate of the late 1990s, users opted for cheaper, though less secure and less comprehensive, cell phones. By the end of the decade, Iridium had declared bankruptcy, shut down operations, and was just waiting for the satellites to start plunging from their orbits around 2007.

The only offer for Iridium's $5 billion system came from an ex-CEO of a nuclear reactor business, Dan Colussy, and it was for a measly $25 million. "It's like picking up a $150,000 Porsche 911 for $750," wrote *USA Today* reporter, Kevin Maney.

The purchase turned into a bonanza. In the wake of September 11, 2001, and then the wars in Afghanistan and Iraq, demand for secure communications in remote locations skyrocketed. New customers included the U.S. and British militaries, as well as reporters in Iraq, who, when traveling with the military have been barred from using less secure systems that are easier to track. The nonprofit organization Operation Call Home has bought time to allow members of the 81st Armor Brigade of the Washington National Guard to communicate with their families at home. Airlines and shipping lines have also signed up.

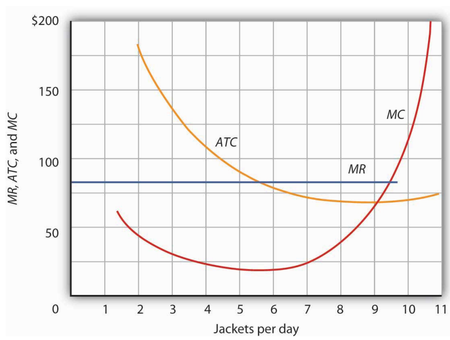
As the new Iridium became unburdened from the debt of the old one and technology improved, the lower fixed and variable costs have contributed to Iridium's revival, but clearly a critical element in the turnaround has been increased demand. The launching of an additional seven spare satellites and other tinkering have extended the life of the system to at least 2014. The firm was temporarily shut down but, with its new owners and new demand for its services, has come roaring back.

Why did Colussy buy Iridium? A top executive in the new firm said that Colussy just found the elimination of the satellites a terrible waste. Perhaps he had some niche uses in mind, as even before September 11, 2001, he had begun to enroll some new customers, such as the Colombian national police, who no doubt found the system useful in the fighting drug lords. But it was in the aftermath of 9/11 that its subscriber list really began to grow and its re-opening was deemed a stroke of genius. Today Iridium's customers include ships at sea (which account for about half of its business), airlines, military uses, and a variety of commercial and humanitarian applications.

Sources: Kevin Maney, "Remember Those 'Iridium's Going to Fail' Jokes? Prepare to Eat Your Hat," USA Today, April 9, 2003: p. 3B. Michael Mecham, "Handheld Comeback: A Resurrected Iridium Counts Aviation, Antiterrorism Among Its Growth Fields," Aviation Week and Space Technology, 161: 9 (September 6, 2004): p. 58. Iridium's webpage can be found at Iridium.com.

ANSWER TO TRY IT! PROBLEM

At a price of $81, Acme's marginal revenue curve is a horizontal line at $81. The firm produces the output at which marginal cost equals marginal revenue; the curves intersect at a quantity of 9 jackets per day. Acme's average total cost at this level of output equals $67, for an economic profit per jacket of $14. Acme's economic profit per day equals about $126.



3. PERFECT COMPETITION IN THE LONG RUN

LEARNING OBJECTIVES

1. **Distinguish between economic profit and accounting profit.**
2. **Explain why in long-run equilibrium in a perfectly competitive industry firms will earn zero economic profit.**
3. **Describe the three possible effects on the costs of the factors of production that expansion or contraction of a perfectly competitive industry may have and illustrate the resulting long-run industry supply curve in each case.**
4. **Explain why under perfection competition output prices will change by less than the change in production cost in the short run, but by the full amount of the change in production cost in the long run.**
5. **Explain the effect of a change in fixed cost on price and output in the short run and in the long run under perfect competition.**

In the long run, a firm is free to adjust all of its inputs. New firms can enter any market; existing firms can leave their markets. We shall see in this section that the model of perfect competition predicts that, at a long-run equilibrium, production takes place at the lowest possible cost per unit and that all economic profits and losses are eliminated.

3.1 Economic Profit and Economic Loss

Economic profits and losses play a crucial role in the model of perfect competition. The existence of economic profits in a particular industry attracts new firms to the industry in the long run. As new firms enter, the supply curve shifts to the right, price falls, and profits fall. Firms continue to enter the industry until economic profits fall to zero. If firms in an industry are experiencing economic losses, some will leave. The supply curve shifts to the left, increasing price and reducing losses. Firms continue to leave until the remaining firms are no longer suffering losses—until economic profits are zero.

Before examining the mechanism through which entry and exit eliminate economic profits and losses, we shall examine an important key to understanding it: the difference between the accounting and economic concepts of profit and loss.

Economic Versus Accounting Concepts of Profit and Loss

Economic profit equals total revenue minus total cost, where cost is measured in the economic sense as opportunity cost. An economic loss (negative economic profit) is incurred if total cost exceeds total revenue.

Accountants include only explicit costs in their computation of total cost. **Explicit costs** include charges that must be paid for factors of production such as labor and capital, together with an estimate of depreciation. Profit computed using only explicit costs is called **accounting profit**. It is the measure of profit firms typically report; firms pay taxes on their accounting profits, and a corporation reporting its profit for a particular period reports its accounting profits. To compute his accounting profits, Mr. Gortari, the radish farmer, would subtract explicit costs, such as charges for labor, equipment, and other supplies, from the revenue he receives.

explicit costs

Charges that must be paid for factors of production such as labor and capital.

accounting profit

Profit computed using only explicit costs.

implicit cost

A cost that is included in the economic concept of opportunity cost but that is not an explicit cost.

Economists recognize costs in addition to the explicit costs listed by accountants. If Mr. Gortari were not growing radishes, he could be doing something else with the land and with his own efforts. Suppose the most valuable alternative use of his land would be to produce carrots, from which Mr. Gortari could earn $250 per month in accounting profits. The income he forgoes by not producing carrots is an opportunity cost of producing radishes. This cost is not explicit; the return Mr. Gortari could get from producing carrots will not appear on a conventional accounting statement of his accounting profit. A cost that is included in the economic concept of opportunity cost, but that is not an explicit cost, is called an **implicit cost**.

The Long Run and Zero Economic Profits

Given our definition of economic profits, we can easily see why, in perfect competition, they must always equal zero in the long run. Suppose there are two industries in the economy, and that firms in Industry A are earning economic profits. By definition, firms in Industry A are earning a return greater than the return available in Industry B. That means that firms in Industry B are earning less than they could in Industry A. Firms in Industry B are experiencing economic losses.

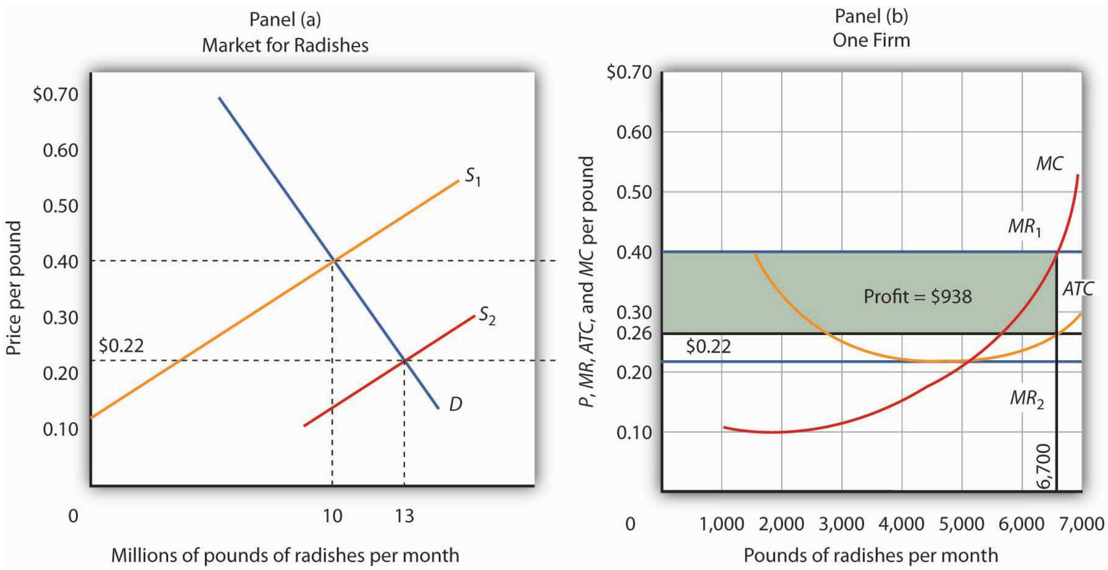
Given easy entry and exit, some firms in Industry B will leave it and enter Industry A to earn the greater profits available there. As they do so, the supply curve in Industry B will shift to the left, increasing prices and profits there. As former Industry B firms enter Industry A, the supply curve in Industry A will shift to the right, lowering profits in A. The process of firms leaving Industry B and entering A will continue until firms in both industries are earning zero economic profit. That suggests an important long-run result: *Economic profits in a system of perfectly competitive markets will, in the long run, be driven to zero in all industries.*

Eliminating Economic Profit: The Role of Entry

The process through which entry will eliminate economic profits in the long run is illustrated in Figure 9.14, which is based on the situation presented in Figure 9.7. The price of radishes is $0.40 per pound. Mr. Gortari's average total cost at an output of 6,700 pounds of radishes per month is $0.26 per pound. Profit per unit is $0.14 ($\$0.40 - \$0.26$). Mr. Gortari thus earns a profit of $938 per month ($= \$0.14 \times 6,700$).

FIGURE 9.14 Eliminating Economic Profits in the Long Run

If firms in an industry are making an economic profit, entry will occur in the long run. In Panel (b), a single firm's profit is shown by the shaded area. Entry continues until firms in the industry are operating at the lowest point on their respective average total cost curves, and economic profits fall to zero.



Profits in the radish industry attract entry in the long run. Panel (a) of Figure 9.14 shows that as firms enter, the supply curve shifts to the right and the price of radishes falls. New firms enter as long as there are economic profits to be made—as long as price exceeds *ATC* in Panel (b). As price falls, marginal revenue falls to MR_2 and the firm reduces the quantity it supplies, moving along the marginal cost (*MC*) curve to the lowest point on the *ATC* curve, at $0.22 per pound and an output of 5,000 pounds per month. Although the output of individual firms falls in response to falling prices, there are now more firms, so industry output rises to 13 million pounds per month in Panel (a).

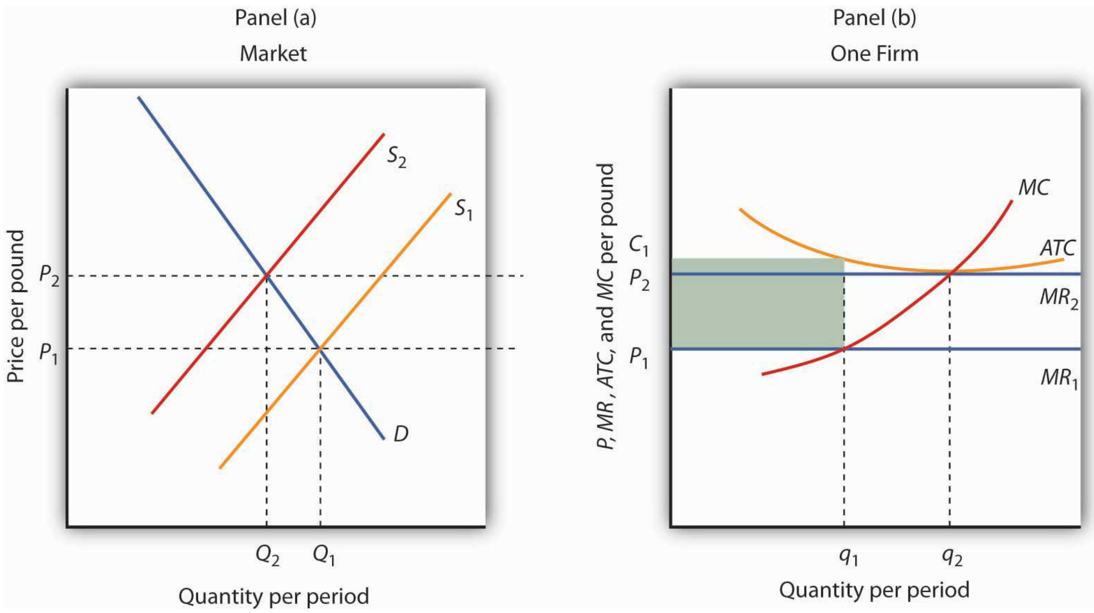
Eliminating Losses: The Role of Exit

Just as entry eliminates economic profits in the long run, exit eliminates economic losses. In Figure 9.15, Panel (a) shows the case of an industry in which the market price P_1 is below *ATC*. In Panel (b), at price P_1 a single firm produces a quantity q_1, assuming it is at least covering its average variable cost. The firm's losses are shown by the shaded rectangle bounded by its average total cost C_1 and price P_1 and by output q_1.

Because firms in the industry are losing money, some will exit. The supply curve in Panel (a) shifts to the left, and it continues shifting as long as firms are suffering losses. Eventually the supply curve shifts all the way to S_2, price rises to P_2, and economic profits return to zero.

FIGURE 9.15 Eliminating Economic Losses in the Long Run

Panel (b) shows that at the initial price P_1, firms in the industry cannot cover average total cost (MR_1 is below ATC). That induces some firms to leave the industry, shifting the supply curve in Panel (a) to S_2, reducing industry output to Q_2 and raising price to P_2. At that price (MR_2), firms earn zero economic profit, and exit from the industry ceases. Panel (b) shows that the firm increases output from q_1 to q_2; total output in the market falls in Panel (a) because there are fewer firms. Notice that in Panel (a) quantity is designated by uppercase Q, while in Panel (b) quantity is designated by lowercase q. This convention is used throughout the text to distinguish between the quantity supplied in the market (Q) and the quantity supplied by a typical firm (q).



Entry, Exit, and Production Costs

constant-cost industry

Industry in which expansion does not affect the prices of factors of production.

In our examination of entry and exit in response to economic profit or loss in a perfectly competitive industry, we assumed that the *ATC* curve of a single firm does not shift as new firms enter or existing firms leave the industry. That is the case when expansion or contraction does not affect prices for the factors of production used by firms in the industry. When expansion of the industry does not affect the prices of factors of production, it is a **constant-cost industry**. In some cases, however, the entry of new firms may affect input prices.

As new firms enter, they add to the demand for the factors of production used by the industry. If the industry is a significant user of those factors, the increase in demand could push up the market price of factors of production for all firms in the industry. If that occurs, then entry into an industry will boost average costs at the same time as it puts downward pressure on price. Long-run equilibrium will still occur at a zero level of economic profit and with firms operating on the lowest point on the *ATC* curve, but that cost curve will be somewhat higher than before entry occurred. Suppose, for example, that an increase in demand for new houses drives prices higher and induces entry. That will increase the demand for workers in the construction industry and is likely to result in higher wages in the industry, driving up costs.

increasing-cost industry

Industry in which the entry of new firms bids up the prices of factors of production and thus increases production costs.

An industry in which the entry of new firms bids up the prices of factors of production and thus increases production costs is called an **increasing-cost industry**. As such an industry expands in the long run, its price will rise.

decreasing-cost industry

Industry in which production costs fall in the long run as firms enter.

Some industries may experience reductions in input prices as they expand with the entry of new firms. That may occur because firms supplying the industry experience economies of scale as they increase production, thus driving input prices down. Expansion may also induce technological changes that lower input costs. That is clearly the case of the computer industry, which has enjoyed falling input costs as it has expanded. An industry in which production costs fall as firms enter in the long run is a **decreasing-cost industry**.

Just as industries may expand with the entry of new firms, they may contract with the exit of existing firms. In a constant-cost industry, exit will not affect the input prices of remaining firms. In an increasing-cost industry, exit will reduce the input prices of remaining firms. And, in a decreasing-cost industry, input prices may rise with the exit of existing firms.

The behavior of production costs as firms in an industry expand or reduce their output has important implications for the **long-run industry supply curve**, a curve that relates the price of a good or service to the quantity produced after all long-run adjustments to a price change have been completed. Every point on a long-run supply curve therefore shows a price and quantity supplied at which firms in the industry are earning zero economic profit. Unlike the short-run market supply curve, the long-run industry supply curve does not hold factor costs and the number of firms unchanged.

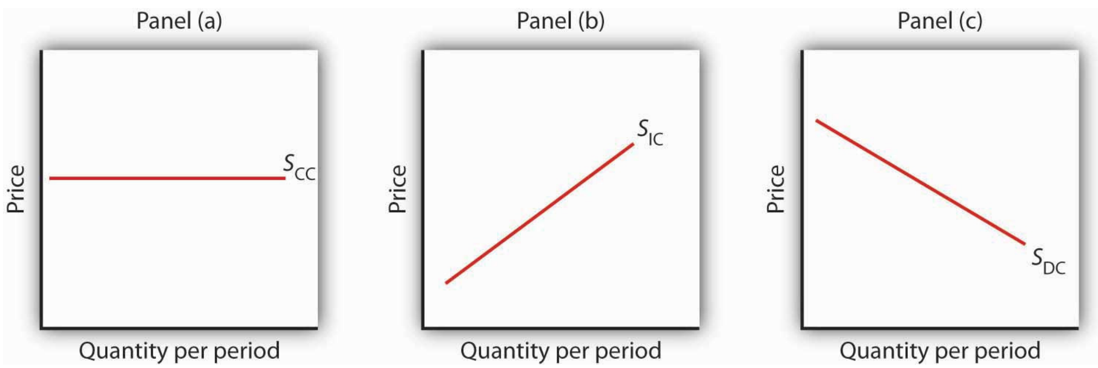
long-run industry supply curve

A curve that relates the price of a good or service to the quantity produced after all long-run adjustments to a price change have been completed.

Figure 9.16 shows three long-run industry supply curves. In Panel (a), S_{CC} is a long-run supply curve for a constant-cost industry. It is horizontal. Neither expansion nor contraction by itself affects market price. In Panel (b), S_{IC} is a long-run supply curve for an increasing-cost industry. It rises as the industry expands. In Panel (c), S_{DC} is a long-run supply curve for a decreasing-cost industry. Its downward slope suggests a falling price as the industry expands.

FIGURE 9.16 Long-Run Supply Curves in Perfect Competition

The long-run supply curve for a constant-cost, perfectly competitive industry is a horizontal line, S_{CC}, shown in Panel (a). The long-run curve for an increasing-cost industry is an upward-sloping curve, S_{IC}, as in Panel (b). The downward-sloping long-run supply curve, S_{DC}, for a decreasing cost industry is given in Panel (c).



3.2 Changes in Demand and in Production Cost

The primary application of the model of perfect competition is in predicting how firms will respond to changes in demand and in production costs. To see how firms respond to a particular change, we determine how the change affects demand or cost conditions and then see how the profit-maximizing solution is affected in the short run and in the long run. Having determined how the profit-maximizing firms of the model would respond, we can then predict firms' responses to similar changes in the real world.

In the examples that follow, we shall assume, for simplicity, that entry or exit do not affect the input prices facing firms in the industry. That is, we assume a constant-cost industry with a horizontal long-run industry supply curve similar to S_{CC} in Figure 9.16. We shall assume that firms are covering their average variable costs, so we can ignore the possibility of shutting down.

Changes in Demand

Changes in demand can occur for a variety of reasons. There may be a change in preferences, incomes, the price of a related good, population, or consumer expectations. A change in demand causes a change in the market price, thus shifting the marginal revenue curves of firms in the industry.

Let us consider the impact of a change in demand for oats. Suppose new evidence suggests that eating oats not only helps to prevent heart disease, but also prevents baldness in males. This will, of course, increase the demand for oats. To assess the impact of this change, we assume that the industry is perfectly competitive and that it is initially in long-run equilibrium at a price of $1.70 per bushel. Economic profits equal zero.

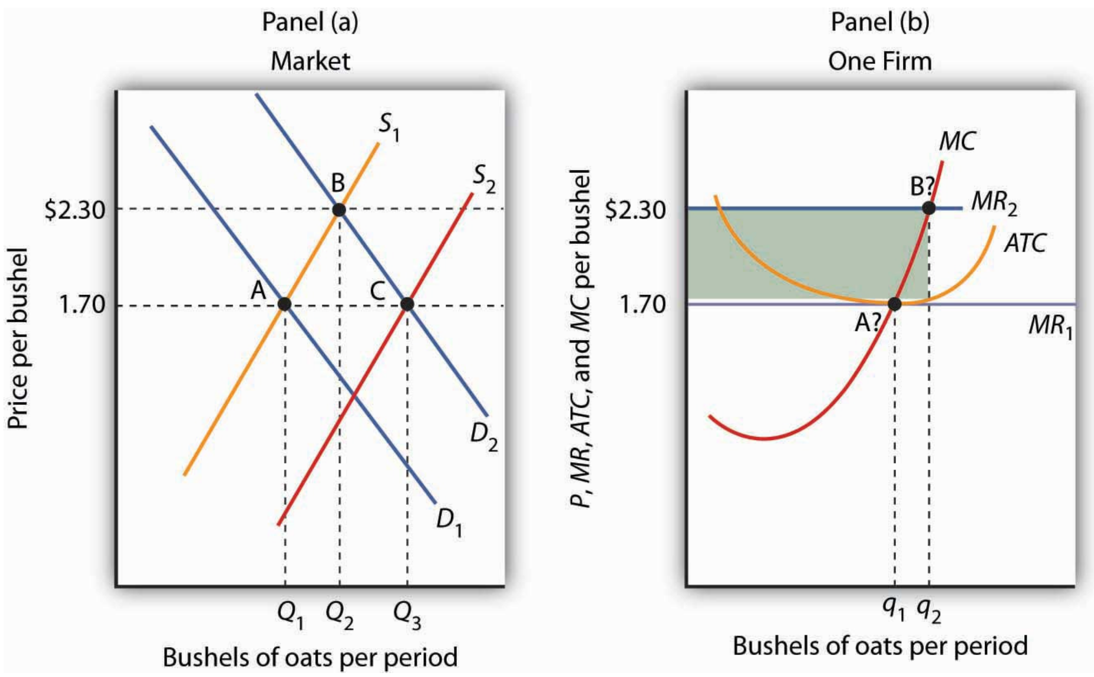
The initial situation is depicted in Figure 9.17. Panel (a) shows that at a price of $1.70, industry output is Q_1 (point A), while Panel (b) shows that the market price constitutes the marginal revenue, MR_1, facing a single firm in the industry. The firm responds to that price by finding the output level at which the MC and MR_1 curves intersect. That implies a level of output q_1 at point A′.

The new medical evidence causes demand to increase to D_2 in Panel (a). That increases the market price to $2.30 (point B), so the marginal revenue curve for a single firm rises to MR_2 in Panel (b). The firm responds by increasing its output to q_2 in the short run (point B′). Notice that the firm's average total cost is slightly higher than its original level of $1.70; that is because of the U shape of the curve. The firm is making an economic profit shown by the shaded rectangle in Panel (b). Other firms in the industry will earn an economic profit as well, which, in the long run, will attract entry by new firms.

New entry will shift the supply curve to the right; entry will continue as long as firms are making an economic profit. The supply curve in Panel (a) shifts to S_2, driving the price down in the long run to the original level of $1.70 per bushel and returning economic profits to zero in long-run equilibrium. A single firm will return to its original level of output, q_1 (point A′) in Panel (b), but because there are more firms in the industry, industry output rises to Q_3 (point C) in Panel (a).

FIGURE 9.17 Short-Run and Long-Run Adjustments to an Increase in Demand

The initial equilibrium price and output are determined in the market for oats by the intersection of demand and supply at point A in Panel (a). An increase in the market demand for oats, from D_1 to D_2 in Panel (a), shifts the equilibrium solution to point B. The price increases in the short run from $1.70 per bushel to $2.30. Industry output rises to Q_2. For a single firm, the increase in price raises marginal revenue from MR_1 to MR_2; the firm responds in the short run by increasing its output to q_2. It earns an economic profit given by the shaded rectangle. In the long run, the opportunity for profit attracts new firms. In a constant-cost industry, the short-run supply curve shifts to S_2; market equilibrium now moves to point C in Panel (a). The market price falls back to $1.70. The firm's demand curve returns to MR_1, and its output falls back to the original level, q_1. Industry output has risen to Q_3 because there are more firms.



A reduction in demand would lead to a reduction in price, shifting each firm's marginal revenue curve downward. Firms would experience economic losses, thus causing exit in the long run and shifting the supply curve to the left. Eventually, the price would rise back to its original level, assuming changes in industry output did not lead to changes in input prices. There would be fewer firms in the industry, but each firm would end up producing the same output as before.

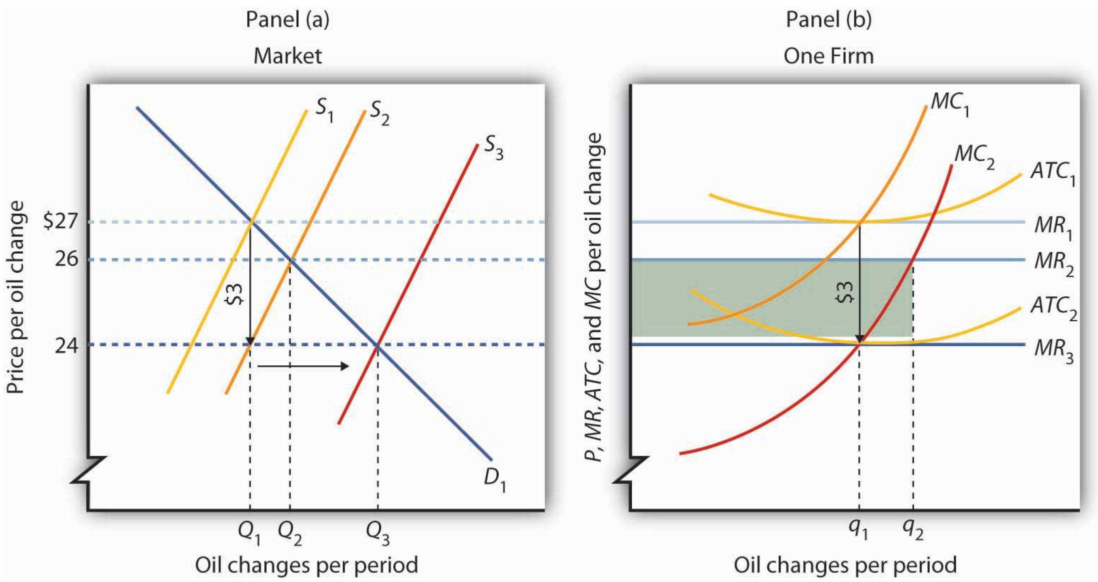
Changes in Production Cost

A firm's costs change if the costs of its inputs change. They also change if the firm is able to take advantage of a change in technology. Changes in production cost shift the *ATC* curve. If a firm's variable costs are affected, its marginal cost curves will shift as well. Any change in marginal cost produces a similar change in industry supply, since it is found by adding up marginal cost curves for individual firms.

Suppose a reduction in the price of oil reduces the cost of producing oil changes for automobiles. We shall assume that the oil-change industry is perfectly competitive and that it is initially in long-run equilibrium at a price of $27 per oil change, as shown in Panel (a) of Figure 9.18. Suppose that the reduction in oil prices reduces the cost of an oil change by $3.

FIGURE 9.18 A Reduction in the Cost of Producing Oil Changes

The initial equilibrium price, \$27, and quantity, Q_1, of automobile oil changes are determined by the intersection of market demand, D_1, and market supply, S_1 in Panel (a). The industry is in long-run equilibrium; a typical firm, shown in Panel (b), earns zero economic profit. A reduction in oil prices reduces the marginal and average total costs of producing an oil change by \$3. The firm's marginal cost curve shifts to MC_2, and its average total cost curve shifts to ATC_2. The short-run industry supply curve shifts down by \$3 to S_2. The market price falls to \$26; the firm increases its output to q_2 and earns an economic profit given by the shaded rectangle. In the long run, the opportunity for profit shifts the industry supply curve to S_3. The price falls to \$24, and the firm reduces its output to the original level, q_1. It now earns zero economic profit once again. Industry output in Panel (a) rises to Q_3 because there are more firms; price has fallen by the full amount of the reduction in production costs.



A reduction in production cost shifts the firm's cost curves down. The firm's average total cost and marginal cost curves shift down, as shown in Panel (b). In Panel (a) the supply curve shifts from S_1 to S_2. The industry supply curve is made up of the marginal cost curves of individual firms; because each of them has shifted downward by \$3, the industry supply curve shifts downward by \$3.

Notice that price in the short run falls to \$26; it does not fall by the \$3 reduction in cost. That is because the supply and demand curves are sloped. While the supply curve shifts downward by \$3, its intersection with the demand curve falls by less than \$3. The firm in Panel (b) responds to the lower price and lower cost by increasing output to q_2, where MC_2 and MR_2 intersect. That leaves firms in the industry with an economic profit; the economic profit for the firm is shown by the shaded rectangle in Panel (b). Profits attract entry in the long run, shifting the supply curve to the right to S_3 in Panel (a) Entry will continue as long as firms are making an economic profit—it will thus continue until the price falls by the full amount of the \$3 reduction in cost. The price falls to \$24, industry output rises to Q_3, and the firm's output returns to its original level, q_1.

An increase in variable costs would shift the average total, average variable, and marginal cost curves upward. It would shift the industry supply curve upward by the same amount. The result in the short run would be an increase in price, but by less than the increase in cost per unit. Firms would experience economic losses, causing exit in the long run. Eventually, price would increase by the full amount of the increase in production cost.

Some cost increases will not affect marginal cost. Suppose, for example, that an annual license fee of \$5,000 is imposed on firms in a particular industry. The fee is a fixed cost; it does not affect marginal cost. Imposing such a fee shifts the average total cost curve upward but causes no change in marginal cost. There is no change in price or output in the short run. Because firms are suffering economic losses, there will be exit in the long run. Prices ultimately rise by enough to cover the cost of the fee, leaving the remaining firms in the industry with zero economic profit.

Price will change to reflect whatever change we observe in production cost. A change in variable cost causes price to change in the short run. In the long run, any change in average total cost changes price by an equal amount.

The message of long-run equilibrium in a competitive market is a profound one. The ultimate beneficiaries of the innovative efforts of firms are consumers. Firms in a perfectly competitive world earn zero profit in the long-run. While firms can earn accounting profits in the long-run, they cannot earn economic profits.

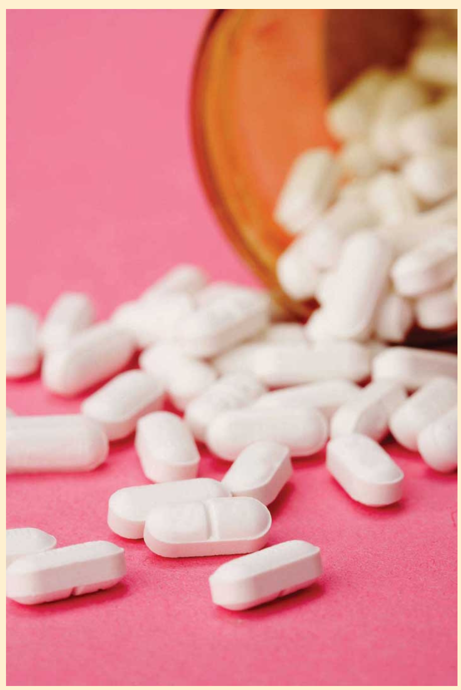
KEY TAKEAWAYS

- The economic concept of profit differs from accounting profit. The accounting concept deals only with explicit costs, while the economic concept of profit incorporates explicit and implicit costs.
- The existence of economic profits attracts entry, economic losses lead to exit, and in long-run equilibrium, firms in a perfectly competitive industry will earn zero economic profit.
- The long-run supply curve in an industry in which expansion does not change input prices (a constant-cost industry) is a horizontal line. The long-run supply curve for an industry in which production costs increase as output rises (an increasing-cost industry) is upward sloping. The long-run supply curve for an industry in which production costs decrease as output rises (a decreasing-cost industry) is downward sloping.
- In a perfectly competitive market in long-run equilibrium, an increase in demand creates economic profit in the short run and induces entry in the long run; a reduction in demand creates economic losses (negative economic profits) in the short run and forces some firms to exit the industry in the long run.
- When production costs change, price will change by less than the change in production cost in the short run. Price will adjust to reflect fully the change in production cost in the long run.
- A change in fixed cost will have no effect on price or output in the short run. It will induce entry or exit in the long run so that price will change by enough to leave firms earning zero economic profit.

TRY IT!

Consider Acme Clothing's situation in the second Try It! in this chapter. Suppose this situation is typical of firms in the jacket market. Explain what will happen in the market for jackets in the long run, assuming nothing happens to the prices of factors of production used by firms in the industry. What will happen to the equilibrium price? What is the equilibrium level of economic profits?

Case in Point: Competition in the Market for Generic Prescription Drugs



Generic prescription drugs are essentially identical substitutes for more expensive brand-name prescription drugs. Since the passage of the Drug Competition and Patent Term Restoration Act of 1984 (commonly referred to as the Hatch-Waxman Act) made it easier for manufacturers to enter the market for generic drugs, the generic drug industry has taken off. Generic drugs represented 19% of the prescription drug industry in 1984 and today represent more than half of the industry. U.S. generic sales were $15 billion in 2002 and soared to $192 billion in 2006. In 2006, the average price of a branded prescription was $111.02 compared to $32.23 for a generic prescription.

A Congressional Budget Office study in the late 1990s showed that entry into the generic drug industry has been the key to this price differential. As shown in the table, when there are one to five manufacturers selling generic copies of a given branded drug, the ratio of the generic price to the branded price is about 60%. With more than 20 competitors, the ratio falls to about 40%.

The generic drug industry is largely characterized by the attributes of a perfectly competitive market. Competitors have good information about the product and sell identical products. The largest generic drug manufacturer in the CBO study had a 16% share of the generic drug manufacturing industry, but most generic manufacturers' sales constituted only 1% to 5% of the market. The 1984 legislation eased entry into this market. And, as the model of perfect competition predicts, entry has driven prices down, benefiting consumers to the tune of tens of billions of dollars each year.

Price Comparison of Generic and Innovator Drugs, by Number of Manufacturers

Number of Generic Manufacturers of a Given Innovator Drug	***Number of Innovator Drugs in Category***	***Avg. Rx Price, All Generic Drugs in Category***	***Avg. Rx Price, All Innovator Drugs in Category***	***Avg. Ratio of the Generic Price to the Innovator Price for Same Drug***
1 to 5	34	$23.40	$37.20	0.61
6 to 10	26	$26.40	$42.60	0.61
11 to 15	29	$20.90	$50.20	0.42
16 to 20	19	$19.90	$45.00	0.46
21 to 24	4	$11.50	$33.90	0.39
Average		$22.40	$43.00	0.53

Sources: Congressional Budget Office, "How Increased Competition from Generic Drugs Has Affected Prices and Returns in the Pharmaceutical Industry," July 1998. Available at www.cbo.gov; "Generic Pharmaceutical Industry Anticipates Double-Digit Growth," PR Newswire, March 17, 2004. Available at www.Prnewswire.com; 2008 Statistical Abstract of the United States, Table 130.

ANSWER TO TRY IT! PROBLEM

The availability of economic profits will attract new firms to the jacket industry in the long run, shifting the market supply curve to the right. Entry will continue until economic profits are eliminated. The price will fall; Acme's marginal revenue curve shifts down. The equilibrium level of economic profits in the long run is zero.

4. REVIEW AND PRACTICE

Summary

The assumptions of the model of perfect competition ensure that every decision maker is a price taker—the interaction of demand and supply in the market determines price. Although most firms in real markets have some control over their prices, the model of perfect competition suggests how changes in demand or in production cost will affect price and output in a wide range of real-world cases.

A firm in perfect competition maximizes profit in the short run by producing an output level at which marginal revenue equals marginal cost, provided marginal revenue is at least as great as the minimum value of average variable cost. For a perfectly competitive firm, marginal revenue equals price and average revenue. This implies that the firm's marginal cost curve is its short-run supply curve for values greater than average variable cost. If price drops below average variable cost, the firm shuts down.

If firms in an industry are earning economic profit, entry by new firms will drive price down until economic profit achieves its long-run equilibrium value of zero. If firms are suffering economic losses, exit by existing firms will continue until price rises to eliminate the losses and economic profits are zero. A long-run equilibrium may be changed by a change in demand or in production cost, which would affect supply. The adjustment to the change in the short run is likely to result in economic profits or losses; these will be eliminated in the long run by entry or by exit.

CONCEPT PROBLEMS

1. Explain how each of the assumptions of perfect competition contributes to the fact that all decision makers in perfect competition are price takers.
2. If the assumptions of perfect competition are not likely to be met in the real world, how can the model be of any use?
3. Explain the difference between marginal revenue, average revenue, and price in perfect competition.
4. Suppose the only way a firm can increase its sales is to lower its price. Is this a perfectly competitive firm? Why or why not?
5. Consider the following goods and services. Which are the most likely to be produced in a perfectly competitive industry? Which are not? Explain why you made the choices you did, relating your answer to the assumptions of the model of perfect competition.
   a. Coca-Cola and Pepsi
   b. Potatoes
   c. Private physicians in your local community
   d. Government bonds and corporate stocks
   e. Taxicabs in Lima, Peru—a city that does not restrict entry or the prices drivers can charge
   f. Oats
6. Explain why an economic profit of zero is acceptable to a firm.
7. Explain why a perfectly competitive firm whose average total cost exceeds the market price may continue to operate in the short run. What about the long run?
8. You have decided to major in biology rather than computer science. A news report suggests that the salaries of computer science majors are increasing. How does this affect the opportunity cost of your choice?
9. Explain how each of the following events would affect the marginal cost curves of firms and thus the supply curve in a perfectly competitive market in the short run.
   a. An increase in wages
   b. A tax of $1 per unit of output imposed on the seller
   c. The introduction of cost-cutting technology
   d. The imposition of an annual license fee of $1,000
10. In a perfectly competitive market, who benefits from an event that lowers production costs for firms?
11. Dry-cleaning establishments generate a considerable amount of air pollution in producing cleaning services. Suppose these firms are allowed to pollute without restriction and that reducing their pollution would add significantly to their production costs. Who benefits from the fact that they pollute the air? Now suppose the government requires them to reduce their pollution. Who will pay for the cleanup? (Assume dry cleaning is a perfectly competitive industry, and answer these questions from a long-run perspective.)
12. The late columnist William F. Buckley, commenting on a strike by the Teamsters Union against UPS in 1997, offered this bit of economic analysis to explain how UPS had succeeded in reducing its average total cost: "UPS has done this by 'economies of scale.' Up to a point (where the marginal cost equals the price of the marginal unit), the larger the business, the less the per-unit cost." Use the concept of economies of scale, together with the information presented in this chapter, to explain the error in Mr. Buckley's statement. [1]
13. Suppose that a perfectly competitive industry is in long-run equilibrium and experiences an increase in production cost. Who will bear the burden of the increase? Is this fair?
14. Economists argue that the ultimate beneficiaries of the efforts of perfectly competitive firms are consumers. In what sense is this the case? Do the owners of perfectly competitive firms derive any long-run benefit from their efforts?
15. Explain carefully why a fixed license fee does not shift a firm's marginal cost curve in the short run. What about the long run?

NUMERICAL PROBLEMS

1. The graph below provides revenue and cost information for a perfectly competitive firm producing paper clips.

Output	Total Revenue	Total Variable Cost	Total Fixed Cost
1	$1,000	$1,500	$500
2	$2,000	$2,000	$500
3	$3,000	$2,600	$500
4	$4,000	$3,900	$500
5	$5,000	$5,000	$500

   a. How much are total fixed costs?
   b. About how much are total variable costs if 5,000 paper clips are produced?
   c. What is the price of a paper clip?
   d. What is the average revenue from producing paper clips?
   e. What is the marginal revenue of producing paper clips?
   f. Over what output range will this firm earn economic profits?
   g. Over what output range will this firm incur economic losses?
   h. What is the slope of the total revenue curve?
   i. What is the slope of the total cost curve at the profit-maximizing number of paper clips per hour?
   j. At about how many paper clips per hour do economic profits seem to be at a maximum?

2. Suppose rocking-chair manufacturing is a perfectly competitive industry in which there are 1,000 identical firms. Each firm's total cost is related to output per day as follows:

Quantity	Total cost	Quantity	Total cost
0	$500	5	$2,200
1	$1,000	6	$2,700
2	$1,300	7	$3,300
3	$1,500	8	$4,400
4	$1,800		

   a. Prepare a table that shows total variable cost, average total cost, and marginal cost at each level of output.
   b. Plot the average total cost, average variable cost, and marginal cost curves for a single firm (remember that values for marginal cost are plotted at the midpoint of the respective intervals).
   c. What is the firm's supply curve? How many chairs would the firm produce at prices of $350, $450, $550, and $650? (In computing quantities, assume that a firm produces a certain number of completed chairs each day; it does not produce fractions of a chair on any one day.)
   d. Suppose the demand curve in the market for rocking chairs is given by the following table:

Price	Quantity of chairs Demanded/day	Price	Quantity of chairs Demanded/day
$650	5,000	$450	7,000
$550	6,000	$350	8,000

   Plot the market demand curve for chairs. Compute and plot the market supply curve, using the information you obtained for a single firm in part (c). What is the equilibrium price? The equilibrium quantity?
   e. Given your solution in part (d), plot the total revenue and total cost curves for a single firm. Does your graph correspond to your solution in part (c)? Explain.

3. The following table shows the total output, total revenue, total variable cost, and total fixed cost of a firm. What level of output should the firm produce? Should it shut down? Should it exit the industry? Explain.

Output	Total revenue	Total variable cost	Total fixed cost
1	$1,000	$1,500	$500
2	$2,000	$2,000	$500
3	$3,000	$2,600	$500
4	$4,000	$3,900	$500
5	$5,000	$5,000	$500

4. Suppose a rise in fuel costs increases the cost of producing oats by $0.50 per bushel. Illustrate graphically how this change will affect the oat market and a single firm in the market in the short run and in the long run.
5. Suppose the demand for car washes in Collegetown falls as a result of a cutback in college enrollment. Show graphically how the price and output for the market and for a single firm will be affected in the short run and in the long run. Assume the market is perfectly competitive and that it is initially in long-run equilibrium at a price of $12 per car wash. Assume also that input prices don't change as the market responds to the change in demand.
6. Suppose that the market for dry-erase pens is perfectly competitive and that the pens cost $1 each. The industry is in long-run equilibrium. Now suppose that an increase in the cost of ink raises the production cost of the pens by $.25 per pen.
    a. Using a graph that shows the market as a whole and a typical firm in this market, illustrate the short run effects of the change.
    b. Is the price likely to rise by $.25? Why or why not?
    c. If it doesn't, are firms likely to continue to operate in the short run? Why or why not?
    d. What is likely to happen in the long run? Illustrate your results with a large, clearly labeled graph.

ENDNOTES

1. William F. Buckley, "Carey Took on 'Greed' as His Battle Cry," *The Gazette*, 22 August 1997, News 7 (a Universal Press Syndicate column).

CHAPTER 10
Monopoly

START UP: SURROUNDED BY MONOPOLIES

If your college or university is like most, you spend a lot of time, and money, dealing with firms that face very little competition. Your campus bookstore is likely to be the only local firm selling the texts that professors require you to read. Your school may have granted an exclusive franchise to a single firm for food service and to another firm for vending machines. A single firm may provide your utilities—electricity, natural gas, and water.

Unlike the individual firms we have previously studied that operate in a competitive market, taking the price, which is determined by demand and supply, as given, in this chapter we investigate the behavior of firms that have their markets all to themselves. As the only suppliers of particular goods or services, they face the downward-sloping market demand curve alone.

We will find that firms that have their markets all to themselves behave in a manner that is in many respects quite different from the behavior of firms in perfect competition. Such firms continue to use the marginal decision rule in maximizing profits, but their freedom to select from the price and quantity combinations given by the market demand curve affects the way in which they apply this rule.

We will show that a monopoly firm is likely to produce less and charge more for what it produces than firms in a competitive industry. As a result, a monopoly solution is likely to be inefficient from society's perspective. We will explore the policy alternatives available to government agencies in dealing with monopoly firms. First, though, we will look at characteristics of monopoly and at conditions that give rise to monopolies in the first place.

1. THE NATURE OF MONOPOLY

LEARNING OBJECTIVES

1. **Define monopoly and the relationship between price setting and monopoly power.**
2. **List and explain the sources of monopoly power and how they can change over time.**
3. **Define what is meant by a natural monopoly.**

Monopoly is at the opposite end of the spectrum of market models from perfect competition. A **monopoly** firm has no rivals. It is the only firm in its industry. There are no close substitutes for the good or service a monopoly produces. Not only does a monopoly firm have the market to itself, but it also need not worry about other firms entering. In the case of monopoly, entry by potential rivals is prohibitively difficult.

A monopoly does not take the market price as given; it determines its own price. It selects from its demand curve the price that corresponds to the quantity the firm has chosen to produce in order to earn the maximum profit possible. The entry of new firms, which eliminates profit in the long run in a competitive market, cannot occur in the monopoly model.

monopoly

A firm that that is the only producer of a good or service for which there are no close substitutes and for which entry by potential rivals is prohibitively difficult.

price setter

A firm that sets or picks price based on its output decision.

monopoly power

The ability to act as a price setter.

A firm that sets or picks price based on its output decision is called a **price setter**. A firm that acts as a price setter possesses **monopoly power**. We shall see in the next chapter that monopolies are not the only firms that have this power; however, the absence of rivals in monopoly gives it much more price-setting power.

As was the case when we discussed perfect competition in the previous chapter, the assumptions of the monopoly model are rather strong. In assuming there is one firm in a market, we assume there are no other firms producing goods or services that could be considered part of the same market as that of the monopoly firm. In assuming blocked entry, we assume, for reasons we will discuss below, that no other firm can enter that market. Such conditions are rare in the real world. As always with models, we make the assumptions that define monopoly in order to simplify our analysis, not to describe the real world. The result is a model that gives us important insights into the nature of the choices of firms and their impact on the economy.

1.1 Sources of Monopoly Power

Why are some markets dominated by single firms? What are the sources of monopoly power? Economists have identified a number of conditions that, individually or in combination, can lead to domination of a market by a single firm and create barriers that prevent the entry of new firms.

barriers to entry

Characteristic of a particular market that block the entry of new firms in a monopoly market.

Barriers to entry are characteristics of a particular market that block new firms from entering it. They include economies of scale, special advantages of location, high sunk costs, a dominant position in the ownership of some of the inputs required to produce the good, and government restrictions. These barriers may be interrelated, making entry that much more formidable. Although these barriers might allow one firm to gain and hold monopoly control over a market, there are often forces at work that can erode this control.

Economies of Scale

Scale economies and diseconomies define the shape of a firm's long-run average cost (*LRAC*) curve as it increases its output. If long-run average cost declines as the level of production increases, a firm is said to experience *economies of scale*.

natural monopoly

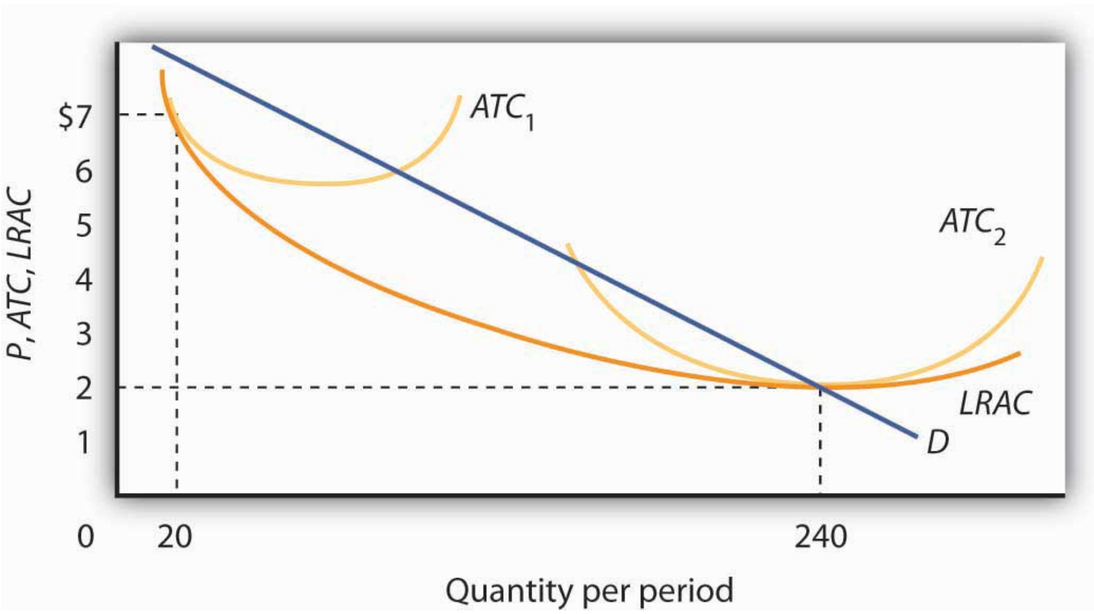
A firm that confronts economies of scale over the entire range of outputs demanded in its industry.

A firm that confronts economies of scale over the entire range of outputs demanded in its industry is a **natural monopoly**. Utilities that distribute electricity, water, and natural gas to some markets are examples. In a natural monopoly, the *LRAC* of any one firm intersects the market demand curve where long-run average costs are falling or are at a minimum. If this is the case, one firm in the industry will expand to exploit the economies of scale available to it. Because this firm will have lower unit costs than its rivals, it can drive them out of the market and gain monopoly control over the industry.

Suppose there are 12 firms, each operating at the scale shown by ATC_1 (average total cost) in Figure 10.1. A firm that expanded its scale of operation to achieve an average total cost curve such as ATC_2 could produce 240 units of output at a lower cost than could the smaller firms producing 20 units each. By cutting its price below the minimum average total cost of the smaller plants, the larger firm could drive the smaller ones out of business. In this situation, the industry demand is not large enough to support more than one firm. If another firm attempted to enter the industry, the natural monopolist would always be able to undersell it.

FIGURE 10.1 Economies of Scale Lead to Natural Monopoly

A firm with falling *LRAC* throughout the range of outputs relevant to existing demand (*D*) will monopolize the industry. Here, one firm operating with a large plant (ATC_2) produces 240 units of output at a lower cost than the $7 cost per unit of the 12 firms operating at a smaller scale (ATC_1), and producing 20 units of output each.



Location

Sometimes monopoly power is the result of location. For example, sellers in markets isolated by distance from their nearest rivals have a degree of monopoly power. The local movie theater in a small town has a monopoly in showing first-run movies. Doctors, dentists, and mechanics in isolated towns may also be monopolists.

Sunk Costs

The greater the cost of establishing a new business in an industry, the more difficult it is to enter that industry. That cost will, in turn, be greater if the outlays required to start a business are unlikely to be recovered if the business should fail.

Suppose, for example, that entry into a particular industry requires extensive advertising to make consumers aware of the new brand. Should the effort fail, there is no way to recover the expenditures for such advertising. An expenditure that has already been made and that cannot be recovered is called a **sunk cost**.

sunk cost

An expenditure that has already been made and that cannot be recovered.

If a substantial fraction of a firm's initial outlays will be lost upon exit from the industry, exit will be costly. Difficulty of exit can make for difficulty of entry. The more firms have to lose from an unsuccessful effort to penetrate a particular market, the less likely they are to try. The potential for high sunk costs could thus contribute to the monopoly power of an established firm by making entry by other firms more difficult.

Restricted Ownership of Raw Materials and Inputs

In very few cases the source of monopoly power is the ownership of strategic inputs. If a particular firm owns all of an input required for the production of a particular good or service, then it could emerge as the only producer of that good or service.

The Aluminum Company of America (ALCOA) gained monopoly power through its ownership of virtually all the bauxite mines in the world (bauxite is the source of aluminum). The International Nickel Company of Canada at one time owned virtually all the world's nickel. De Beers acquired rights to nearly all the world's diamond production, giving it enormous power in the market for diamonds. With new diamond supplies in Canada, Australia, and Russia being developed and sold independently of DeBeers, however, this power has declined, and today DeBeers controls a substantially smaller percentage of the world's supply.

Government Restrictions

Another important basis for monopoly power consists of special privileges granted to some business firms by government agencies. State and local governments have commonly assigned exclusive franchises—rights to conduct business in a specific market—to taxi and bus companies, to cable television companies, and to providers of telephone services, electricity, natural gas, and water, although the trend in recent years has been to encourage competition for many of these services. Governments might also regulate entry into an industry or a profession through licensing and certification requirements. Governments also provide patent protection to inventors of new products or production methods in order to encourage innovation; these patents may afford their holders a degree of monopoly power during the 17-year life of the patent.

Patents can take on extra importance when network effects are present. **Network effects** arise in situations where products become more useful the larger the number of users of the product. For example, one advantage of using the Windows computer operating system is that so many other people use it. That has advantages in terms of sharing files and other information.

network effects

Situations where products become more useful the larger the number of users of the product.

KEY TAKEAWAYS

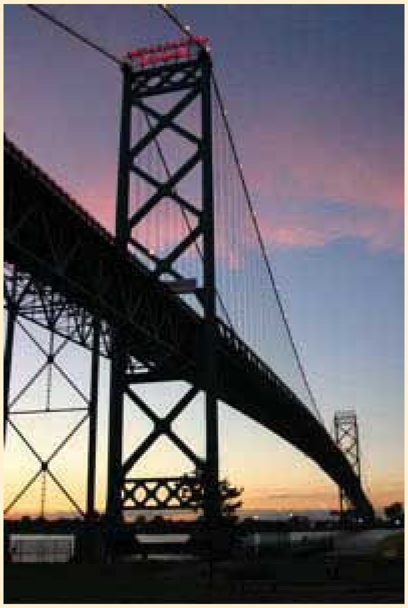
- An industry with a single firm, in which entry is blocked, is called a monopoly.
- A firm that sets or picks price depending on its output decision is called a price setter. A price setter possesses monopoly power.
- The sources of monopoly power include economies of scale, locational advantages, high sunk costs associated with entry, restricted ownership of key inputs, and government restrictions, such as exclusive franchises, licensing and certification requirements, and patents.
- A firm that confronts economies of scale over the entire range of output demanded in an industry is a natural monopoly.

TRY IT!

What is the source of monopoly power—if any—in each of the following situations?

1. The U.S. Food and Drug Administration granted Burroughs Wellcome exclusive rights until 2005 to manufacture and distribute AZT, a drug used in the treatment of AIDS.
2. John and Mary Doe run the only shoe repair shop in town.
3. One utility company distributes residential electricity in your town.
4. The widespread use of automatic teller machines (ATMs) has proven a boon to Diebold, the principal manufacturer of the machines.

Case in Point: The Ambassador Bridge Fights to Maintain Its Monopoly



© 2010 Jupiterimages Corporation

Matty Moroun was quietly enjoying his monopoly power. He is the owner of the 75-year-old Ambassador Bridge, a suspension bridge that is the only connection between Detroit, Michigan and Windsor, Ontario. He purchased the bridge from Warren Buffet in 1974 for $30 million. Forbes estimates that it is worth more than $500 million today. Mr. Moroun now oversees the artery over which $100 billion of goods—one quarter of U.S. trade with Canada and 40% of all truck shipments from the U.S.—make their way between the two countries.

Despite complaints of high and rising tolls—he has more than doubled fares for cars and tripled fares for trucks—Mr. Moroun has so far held on. Kenneth Davies, a lawyer who often battles Mr. Moroun in court, is a grudging admirer. "He's very intelligent and very aggressive. His avarice and greed are just American capitalism at work," he told Forbes.

What are the sources of his monopoly power? With the closest alternative bridge across the Detroit River two hours away, location is a big plus. In addition, the cost of creating a new transportation link is high. A group that is considering converting an old train tunnel to truck use and boring a new train tunnel some distance away is facing a $600 million price tag for the project. In addition to having entry by potential competitors blocked , he has a status not shared by most other monopolists. The Michigan Supreme Court ruled in 2008 that the city of Detroit cannot regulate his business because of the bridge's international nature. Canadian courts have barred any effort by Canadian authorities to regulate him. He will not even allow inspectors from the government of the United States to set foot on his bridge.

Increased security since 9/11 has caused delays, but Mr. Moroun has eased these by increasing his own spending on security to $50,000 a week and by building additional inspection stations and gifting them to the U.S. inspection agency, the General Services Administration. Even a monopolist understands the importance of keeping his customers content!

Because of the terrorist attacks on 9/11 and the concern about vulnerability and security, calls to deal with this monopoly have increased. Some people argue that the government should buy what is the most important single international arterial in North America, while others have called for more regulatory oversight. Canadian groups are exploring the development of alternative means of bringing traffic between the United States and Canada. Time will tell whether Mr. Moroun can hold onto what *Forbes* writers Stephane Fitch and Joann Muller dubbed "the best monopoly you never heard of."

Sources: Stephane Fitch and Joann Muller, "The Troll Under the Bridge," Forbes 174:10 (November 15 2004): 134–139; John Gallagher, "Plan Uses Parkway to Ease Ambassador Bridge Traffic," Detroit Free Press, May 1, 2008; and "State Supreme Court Sides with Ambassador Bridge in Dispute," Detroit News, May 7, 2008.

ANSWERS TO TRY IT! PROBLEMS

1. The government's grant of an exclusive franchise to the drug gave the firm monopoly power.
2. While John and Mary have the only shop in town, this is an easy entry business. Further, there may be competitors in the nearby town. John and Mary probably have monopoly power, but they do not have a monopoly.
3. Natural monopoly
4. Patent with strong network effects

2. THE MONOPOLY MODEL

LEARNING OBJECTIVES

1. **Explain the relationship between price and marginal revenue when a firm faces a downward-sloping demand curve.**
2. **Explain the relationship between marginal revenue and elasticity along a linear demand curve.**
3. **Apply the marginal decision rule to explain how a monopoly maximizes profit.**

Analyzing choices is a more complex challenge for a monopoly firm than for a perfectly competitive firm. After all, a competitive firm takes the market price as given and determines its profit-maximizing output. Because a monopoly has its market all to itself, it can determine not only its output but its price as well. What kinds of price and output choices will such a firm make?

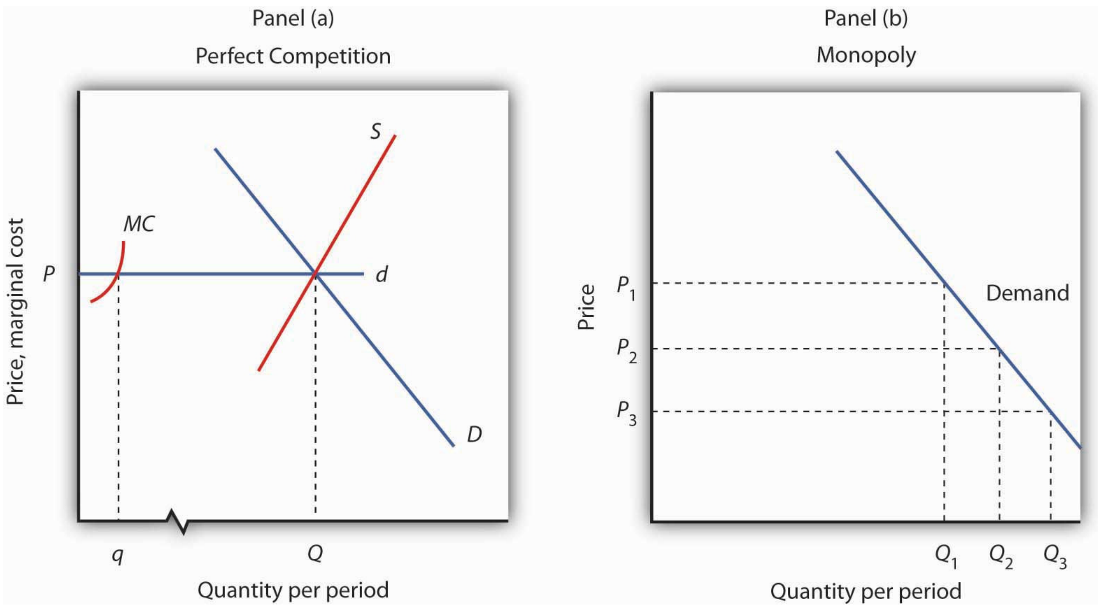
We will answer that question in the context of the marginal decision rule: a firm will produce additional units of a good until marginal revenue equals marginal cost. To apply that rule to a monopoly firm, we must first investigate the special relationship between demand and marginal revenue for a monopoly.

2.1 Monopoly and Market Demand

Because a monopoly firm has its market all to itself, it faces the market demand curve. Figure 10.3 compares the demand situations faced by a monopoly and a perfectly competitive firm. In Panel (a), the equilibrium price for a perfectly competitive firm is determined by the intersection of the demand and supply curves. The market supply curve is found simply by summing the supply curves of individual firms. Those, in turn, consist of the portions of marginal cost curves that lie above the average variable cost curves. The marginal cost curve, *MC*, for a single firm is illustrated. Notice the break in the horizontal axis indicating that the quantity produced by a single firm is a trivially small fraction of the whole. In the perfectly competitive model, one firm has *nothing* to do with the determination of the market price. Each firm in a perfectly competitive industry faces a horizontal demand curve defined by the market price.

FIGURE 10.3 Perfect Competition Versus Monopoly

Panel (a) shows the determination of equilibrium price and output in a perfectly competitive market. A typical firm with marginal cost curve *MC* is a price taker, choosing to produce quantity *q* at the equilibrium price *P*. In Panel (b) a monopoly faces a downward-sloping market demand curve. As a profit maximizer, it determines its profit-maximizing output. Once it determines that quantity, however, the price at which it can sell that output is found from the demand curve. The monopoly firm can sell additional units only by lowering price. The perfectly competitive firm, by contrast, can sell any quantity it wants at the market price.



Contrast the situation shown in Panel (a) with the one faced by the monopoly firm in Panel (b). Because it is the only supplier in the industry, the monopolist faces the downward-sloping market demand curve alone. It may choose to produce any quantity. But, unlike the perfectly competitive firm, which can sell all it wants at the going market price, a monopolist can sell a greater quantity only by cutting its price.

Suppose, for example, that a monopoly firm can sell quantity Q_1 units at a price P_1 in Panel (b). If it wants to increase its output to Q_2 units—and sell that quantity—it must reduce its price to P_2. To sell quantity Q_3 it would have to reduce the price to P_3. The monopoly firm may choose its price and output, but it is restricted to a combination of price and output that lies on the demand curve. It could not, for example, charge price P_1 and sell quantity Q_3. To be a price setter, a firm must face a downward-sloping demand curve.

2.2 Total Revenue and Price Elasticity

A firm's elasticity of demand with respect to price has important implications for assessing the impact of a price change on total revenue. Also, the price elasticity of demand can be different at different points on a firm's demand curve. In this section, we shall see why a monopoly firm will always select a price in the elastic region of its demand curve.

Suppose the demand curve facing a monopoly firm is given by Equation 10.1, where *Q* is the quantity demanded per unit of time and *P* is the price per unit:

EQUATION 10.1

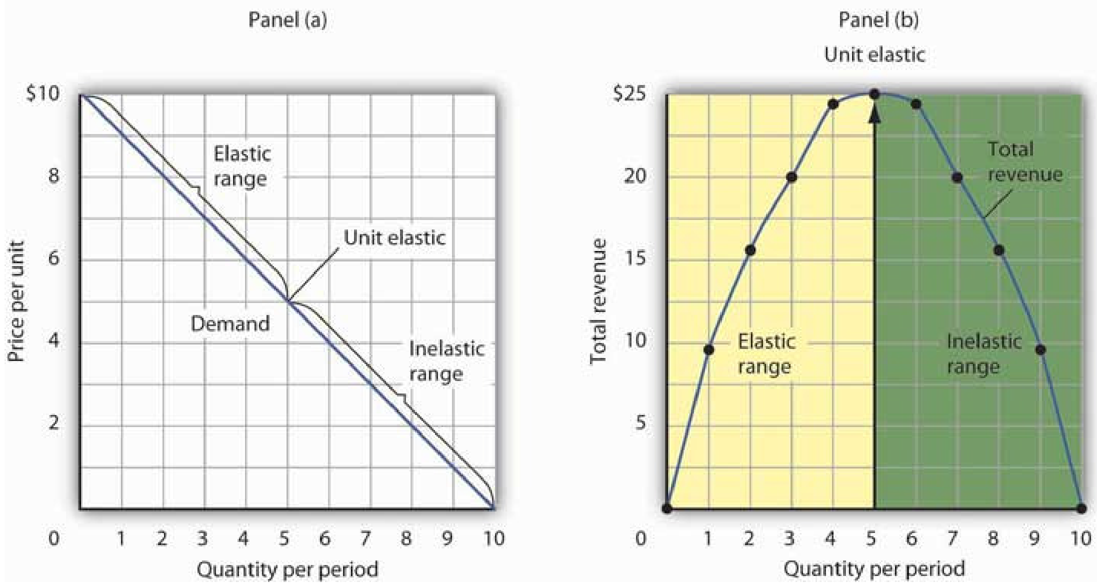
$$Q = 10 - P$$

This demand equation implies the demand schedule shown in Figure 10.4. Total revenue for each quantity equals the quantity times the price at which that quantity is demanded. The monopoly firm's total revenue curve is given in Panel (b). Because a monopolist must cut the price of every unit in order to increase sales, total revenue does not always increase as output rises. In this case, total revenue reaches a maximum of $25 when 5 units are sold. Beyond 5 units, total revenue begins to decline.

FIGURE 10.4 Demand, Elasticity, and Total Revenue

Suppose a monopolist faces the downward-sloping demand curve shown in Panel (a). In order to increase the quantity sold, it must cut the price. Total revenue is found by multiplying the price and quantity sold at each price. Total revenue, plotted in Panel (b), is maximized at $25, when the quantity sold is 5 units and the price is $5. At that point on the demand curve, the price elasticity of demand equals −1.

Price	$10	9	8	7	6	5	4	3	2	1	0
Quantity	0	1	2	3	4	5	6	7	8	9	10
Total revenue	$ 0	9	16	21	24	25	24	21	16	9	0



The demand curve in Panel (a) of Figure 10.4 shows ranges of values of the price elasticity of demand. We have learned that price elasticity varies along a linear demand curve in a special way: Demand is price elastic at points in the upper half of the demand curve and price inelastic in the lower half of the demand curve. If demand is price elastic, a price reduction increases total revenue. To sell an additional unit, a monopoly firm must lower its price. The sale of one more unit will increase revenue because the percentage increase in the quantity demanded exceeds the percentage decrease in the price. The elastic range of the demand curve corresponds to the range over which the total revenue curve is rising in Panel (b) of Figure 10.4.

If demand is price inelastic, a price reduction reduces total revenue because the percentage increase in the quantity demanded is less than the percentage decrease in the price. Total revenue falls as the firm sells additional units over the inelastic range of the demand curve. The downward-sloping portion of the total revenue curve in Panel (b) corresponds to the inelastic range of the demand curve.

Finally, recall that the midpoint of a linear demand curve is the point at which demand becomes unit price elastic. That point on the total revenue curve in Panel (b) corresponds to the point at which total revenue reaches a maximum.

The relationship among price elasticity, demand, and total revenue has an important implication for the selection of the profit-maximizing price and output: A monopoly firm will never choose a price and output in the inelastic range of the demand curve. Suppose, for example, that the monopoly firm represented in Figure 10.4 is charging $3 and selling 7 units. Its total revenue is thus $21. Because this combination is in the inelastic portion of the demand curve, the firm could increase its total revenue by raising its price. It could, at the same time, reduce its total cost. Raising price means reducing output; a reduction in output would reduce total cost. If the firm is operating in the inelastic range of its demand curve, then it is not maximizing profits. The firm could earn a higher profit by raising price and reducing output. It will continue to raise its price until it is in the elastic portion of its demand curve. A profit-maximizing monopoly firm will therefore select a price and output combination in the elastic range of its demand curve.

Of course, the firm could choose a point at which demand is unit price elastic. At that point, total revenue is maximized. But the firm seeks to maximize profit, not total revenue. A solution that maximizes total revenue will not maximize profit unless marginal cost is zero.

2.3 Demand and Marginal Revenue

In the perfectly competitive case, the additional revenue a firm gains from selling an additional unit—its marginal revenue—is equal to the market price. The firm's demand curve, which is a horizontal line at the market price, is also its marginal revenue curve. But a monopoly firm can sell an additional unit only by lowering the price. That fact complicates the relationship between the monopoly's demand curve and its marginal revenue.

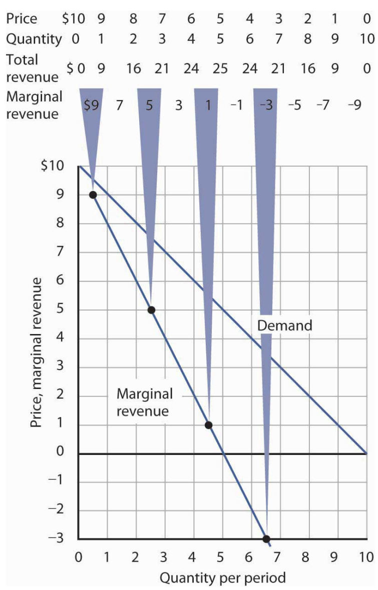
Suppose the firm in Figure 10.4 sells 2 units at a price of $8 per unit. Its total revenue is $16. Now it wants to sell a third unit and wants to know the marginal revenue of that unit. To sell 3 units rather than 2, the firm must lower its price to $7 per unit. Total revenue rises to $21. The marginal revenue of the third unit is thus $5. But the *price* at which the firm sells 3 units is $7. Marginal revenue is less than price.

To see why the marginal revenue of the third unit is less than its price, we need to examine more carefully how the sale of that unit affects the firm's revenues. The firm brings in $7 from the sale of the third unit. But selling the third unit required the firm to charge a price of $7 instead of the $8 the firm was charging for 2 units. Now the firm receives less for the first 2 units. The marginal revenue of the third unit is the $7 the firm receives for that unit *minus* the $1 reduction in revenue for each of the first two units. The marginal revenue of the third unit is thus $5. (In this chapter we assume that the monopoly firm sells all units of output at the same price. In the next chapter, we will look at cases in which firms charge different prices to different customers.)

Marginal revenue is less than price for the monopoly firm. Figure 10.5 shows the relationship between demand and marginal revenue, based on the demand curve introduced in Figure 10.4. As always, we follow the convention of plotting marginal values at the midpoints of the intervals.

FIGURE 10.5 Demand and Marginal Revenue

The marginal revenue curve for the monopoly firm lies below its demand curve. It shows the additional revenue gained from selling an additional unit. Notice that, as always, marginal values are plotted at the midpoints of the respective intervals.



When the demand curve is linear, as in Figure 10.5, the marginal revenue curve can be placed according to the following rules: the marginal revenue curve is always below the demand curve and the marginal revenue curve will bisect any horizontal line drawn between the vertical axis and the demand curve. To put it another way, the marginal revenue curve will be twice as steep as the demand curve. The demand curve in Figure 10.5 is given by the equation $Q = 10 - P$, which can be written $P = 10 - Q$. The marginal revenue curve is given by $P = 10 - 2Q$, which is twice as steep as the demand curve.

The marginal revenue and demand curves in Figure 10.5 follow these rules. The marginal revenue curve lies below the demand curve, and it bisects any horizontal line drawn from the vertical axis to the demand curve. At a price of $6, for example, the quantity demanded is 4. The marginal revenue curve passes through 2 units at this price. At a price of 0, the quantity demanded is 10; the marginal revenue curve passes through 5 units at this point.

Just as there is a relationship between the firm's demand curve and the price elasticity of demand, there is a relationship between its marginal revenue curve and elasticity. Where marginal revenue is positive, demand is price elastic. Where marginal revenue is negative, demand is price inelastic. Where marginal revenue is zero, demand is unit price elastic.

When marginal revenue is …	then demand is …
positive,	price elastic.
negative,	price inelastic.
zero,	unit price elastic.

A firm would not produce an additional unit of output with negative marginal revenue. And, assuming that the production of an additional unit has some cost, a firm would not produce the extra unit if it has zero marginal revenue. Because a monopoly firm will generally operate where marginal revenue is positive, we see once again that it will operate in the elastic range of its demand curve.

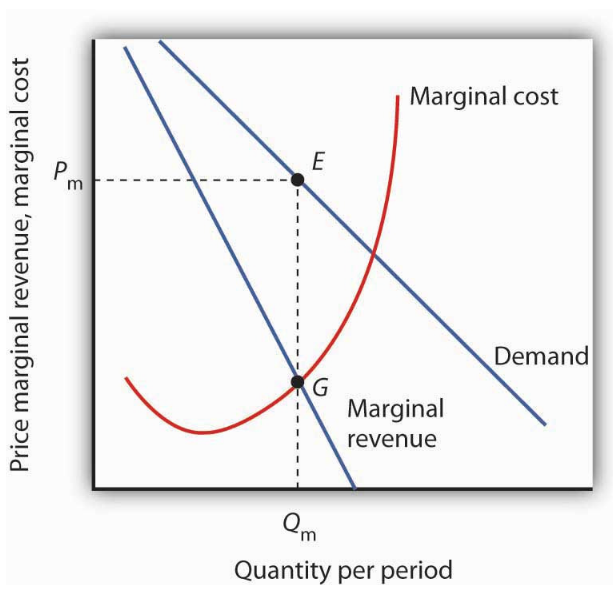
2.4 Monopoly Equilibrium: Applying the Marginal Decision Rule

Profit-maximizing behavior is always based on the marginal decision rule: Additional units of a good should be produced as long as the marginal revenue of an additional unit exceeds the marginal cost. The maximizing solution occurs where marginal revenue equals marginal cost. As always, firms seek to maximize economic profit, and costs are measured in the economic sense of opportunity cost.

Figure 10.6 shows a demand curve and an associated marginal revenue curve facing a monopoly firm. The marginal cost curve is like those we derived earlier; it falls over the range of output in which the firm experiences increasing marginal returns, then rises as the firm experiences diminishing marginal returns.

FIGURE 10.6 The Monopoly Solution

The monopoly firm maximizes profit by producing an output Q_m at point G, where the marginal revenue and marginal cost curves intersect. It sells this output at price P_m.



To determine the profit-maximizing output, we note the quantity at which the firm's marginal revenue and marginal cost curves intersect (Q_m in Figure 10.6). We read up from Q_m to the demand curve to find the price P_m at which the firm can sell Q_m units per period. The profit-maximizing price and output are given by point E on the demand curve.

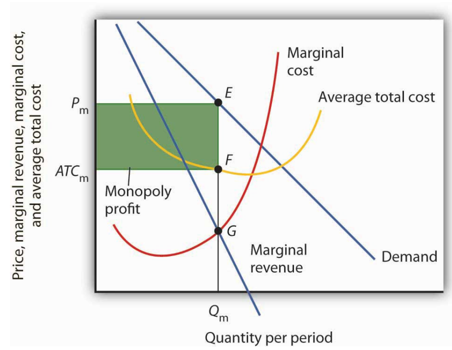
Thus we can determine a monopoly firm's profit-maximizing price and output by following three steps:

1. Determine the demand, marginal revenue, and marginal cost curves.
2. Select the output level at which the marginal revenue and marginal cost curves intersect.
3. Determine from the demand curve the price at which that output can be sold.

Once we have determined the monopoly firm's price and output, we can determine its economic profit by adding the firm's average total cost curve to the graph showing demand, marginal revenue, and marginal cost, as shown in Figure 10.7. The average total cost (ATC) at an output of Q_m units is ATC_m. The firm's profit per unit is thus P_m - ATC_m. Total profit is found by multiplying the firm's output, Q_m, by profit per unit, so total profit equals $Q_m(P_m - ATC_m)$—the area of the shaded rectangle in Figure 10.7.

FIGURE 10.7 Computing Monopoly Profit

A monopoly firm's profit per unit is the difference between price and average total cost. Total profit equals profit per unit times the quantity produced. Total profit is given by the area of the shaded rectangle ATC_mP_mEF.



Heads Up!

Dispelling Myths About Monopoly

Three common misconceptions about monopoly are:

1. Because there are no rivals selling the products of monopoly firms, they can charge whatever they want.
2. Monopolists will charge whatever the market will bear.
3. Because monopoly firms have the market to themselves, they are guaranteed huge profits.

As Figure 10.6 shows, once the monopoly firm decides on the number of units of output that will maximize profit, the price at which it can sell that many units is found by "reading off" the demand curve the price associated with that many units. If it tries to sell Q_m units of output for more than P_m, some of its output will go unsold. The monopoly firm can set its price, but is restricted to price and output combinations that lie on its demand curve. It cannot just "charge whatever it wants." And if it charges "all the market will bear," it will sell either 0 or, at most, 1 unit of output.

Neither is the monopoly firm guaranteed a profit. Consider Figure 10.7. Suppose the average total cost curve, instead of lying below the demand curve for some output levels as shown, were instead everywhere above the demand curve. In that case, the monopoly will incur losses no matter what price it chooses, since average total cost will always be greater than any price it might charge. As is the case for perfect competition, the monopoly firm can keep producing in the short run so long as price exceeds average variable cost. In the long run, it will stay in business only if it can cover all of its costs.

KEY TAKEAWAYS

- If a firm faces a downward-sloping demand curve, marginal revenue is less than price.
- Marginal revenue is positive in the elastic range of a demand curve, negative in the inelastic range, and zero where demand is unit price elastic.
- If a monopoly firm faces a linear demand curve, its marginal revenue curve is also linear, lies below the demand curve, and bisects any horizontal line drawn from the vertical axis to the demand curve.
- To maximize profit or minimize losses, a monopoly firm produces the quantity at which marginal cost equals marginal revenue. Its price is given by the point on the demand curve that corresponds to this quantity.

TRY IT!

The Troll Road Company is considering building a toll road. It estimates that its linear demand curve is as shown below. Assume that the fixed cost of the road is $0.5 million per year. Maintenance costs, which are the only other costs of the road, are also given in the table.

Tolls per trip	$1.00	0.90	0.80	0.70	0.60	0.50
Number of trips per year (in millions)	1	2	3	4	5	6
Maintenance cost per year (in millions)	$0.7	1.2	1.8	2.9	4.2	6.0

1. Using the midpoint convention, compute the profit-maximizing level of output.
2. Using the midpoint convention, what price will the company charge?
3. What is marginal revenue at the profit-maximizing output level? How does marginal revenue compare to price?

Case in Point: Profit-Maximizing Hockey Teams

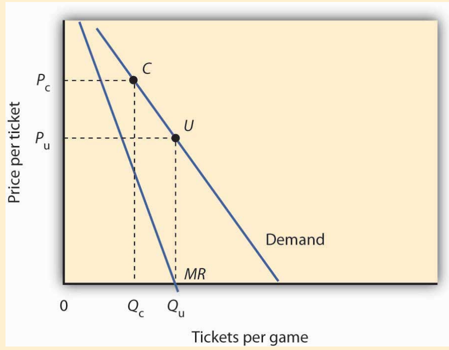
Love of the game? Love of the city? Are those the factors that influence owners of professional sports teams in setting admissions prices? Four economists at the University of Vancouver have what they think is the answer for one group of teams: professional hockey teams set admission prices at levels that maximize their profits. They regard hockey teams as monopoly firms and use the monopoly model to examine the team's behavior.

The economists, Donald G. Ferguson, Kenneth G. Stewart, John Colin H. Jones, and Andre Le Dressay, used data from three seasons to estimate demand and marginal revenue curves facing each team in the National Hockey League. They found that demand for a team's tickets is affected by population and income in the team's home city, the team's standing in the National Hockey League, and the number of superstars on the team.

Because a sports team's costs do not vary significantly with the number of fans who attend a given game, the economists assumed that marginal cost is zero. The profit-maximizing number of seats sold per game is thus the quantity at which marginal revenue is zero, provided a team's stadium is large enough to hold that quantity of fans. This unconstrained quantity is labeled Q_u, with a corresponding price P_u in the graph.

Stadium size and the demand curve facing a team might prevent the team from selling the profit-maximizing quantity of tickets. If its stadium holds only Q_c fans, for example, the team will sell that many tickets at price P_c; its marginal revenue is positive at that quantity. Economic theory thus predicts that the marginal revenue for teams that consistently sell out their games will be positive, and the marginal revenue for other teams will be zero.

The economists' statistical results were consistent with the theory. They found that teams that don't typically sell out their games operate at a quantity at which marginal revenue is about zero, and that teams with sellouts have positive marginal revenue. "It's clear that these teams are very sophisticated in their use of pricing to maximize profits," Mr. Ferguson said.



Sources: Donald G. Ferguson et al., "The Pricing of Sports Events: Do Teams Maximize Profit?" Journal of Industrial Economics 39(3) (March 1991): 297–310 and personal interview.

ANSWER TO TRY IT! PROBLEM

Maintenance costs constitute the variable costs associated with building the road. In order to answer the first four parts of the question, you will need to compute total revenue, marginal revenue, and marginal cost, as shown at right:

1. Using the "midpoint" convention, the profit-maximizing level of output is 2.5 million trips per year. With that number of trips, marginal revenue ($0.60) equals marginal cost ($0.60).
2. Again, we use the "midpoint" convention. The company will charge a toll of $0.85.
3. The marginal revenue is $0.60, which is less than the $0.85 toll (price).

Price per trip	$1.00		0.90		0.80		0.70		0.60		0.50
Number of trips per year (in millions)	1		2		3		4		5		6
Total variable costs per year (in millions)	$0.7		1.2		1.8		2.9		4.2		6.0
Total revenue (in millions)	$1.00		1.80		2.40		2.80		3.00		3.00
Marginal revenue		$0.80		0.60		0.40		0.20		0.00	
Marginal cost		$0.50		0.60		1.10		1.30		1.80	

3. ASSESSING MONOPOLY

LEARNING OBJECTIVES

1. **Explain and illustrate that a monopoly firm produces an output that is less than the efficient level and why this results in a deadweight loss to society.**
2. **Explain and illustrate how the higher price that a monopoly charges, compared to an otherwise identical perfectly competitive firm, transfers part of consumer surplus to the monopolist and raises questions of equity.**
3. **Considering both advantages and disadvantages, discuss the potential effects that a monopoly may have on consumer choices, price, quality of products, and technological innovations.**
4. **Discuss the public policy responses to monopoly.**

We have seen that for monopolies pursuing profit maximization, the outcome differs from the case of perfect competition. Does this matter to society? In this section, we will focus on the differences that stem from market structure and assess their implications.

3.1 Efficiency, Equity, and Concentration of Power

A monopoly firm determines its output by setting marginal cost equal to marginal revenue. It then charges the price at which it can sell that output, a price determined by the demand curve. That price exceeds marginal revenue; it therefore exceeds marginal cost as well. That contrasts with the case in perfect competition, in which price and marginal cost are equal. The higher price charged by a monopoly firm may allow it a profit—in large part at the expense of consumers, whose reduced options may give them little say in the matter. The monopoly solution thus raises problems of efficiency, equity, and the concentration of power.

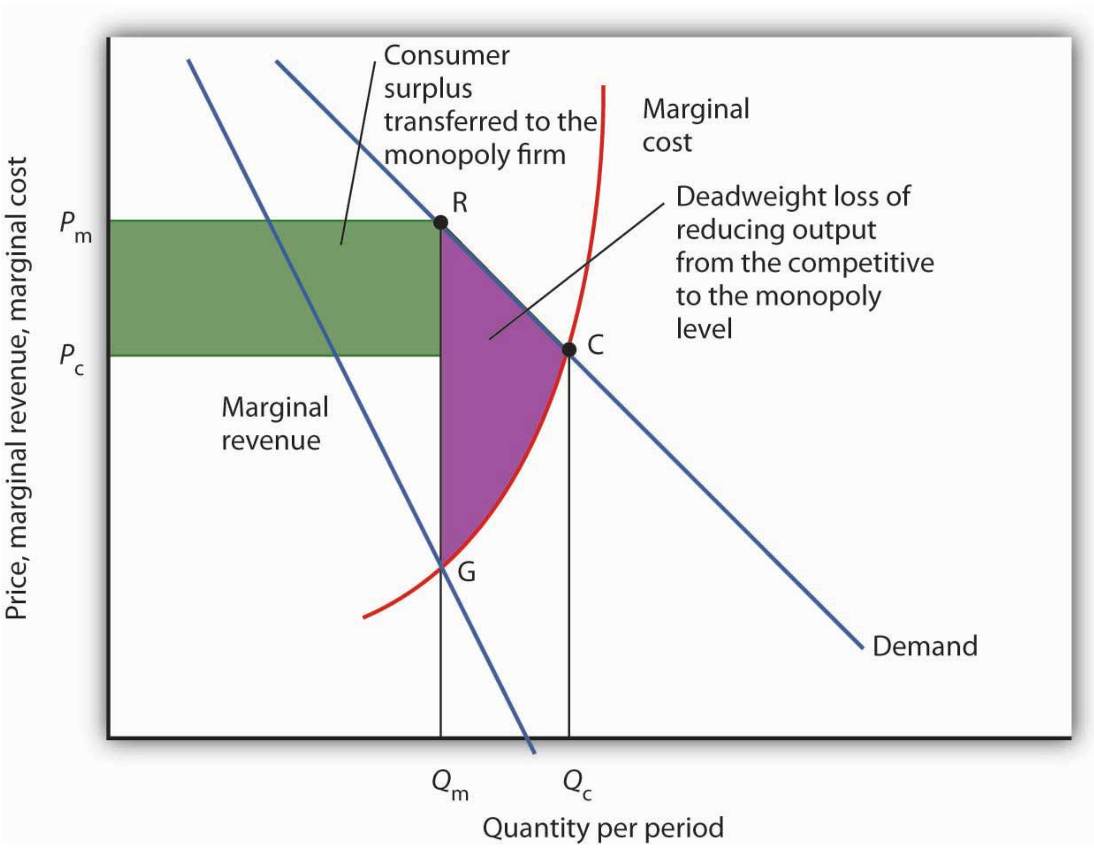
Monopoly and Efficiency

The fact that price in monopoly exceeds marginal cost suggests that the monopoly solution violates the basic condition for economic efficiency, that the price system must confront decision makers with all of the costs and all of the benefits of their choices. Efficiency requires that consumers confront prices that equal marginal costs. Because a monopoly firm charges a price greater than marginal cost, consumers will consume less of the monopoly's good or service than is economically efficient.

To contrast the efficiency of the perfectly competitive outcome with the inefficiency of the monopoly outcome, imagine a perfectly competitive industry whose solution is depicted in Figure 10.11. The short-run industry supply curve is the summation of individual marginal cost curves; it may be regarded as the marginal cost curve for the industry. A perfectly competitive industry achieves equilibrium at point C, at price P_c and quantity Q_c.

FIGURE 10.11 **Perfect Competition, Monopoly, and Efficiency**

Given market demand and marginal revenue, we can compare the behavior of a monopoly to that of a perfectly competitive industry. The marginal cost curve may be thought of as the supply curve of a perfectly competitive industry. The perfectly competitive industry produces quantity Q_c and sells the output at price P_c. The monopolist restricts output to Q_m and raises the price to P_m. Reorganizing a perfectly competitive industry as a monopoly results in a deadweight loss to society given by the shaded area GRC. It also transfers a portion of the consumer surplus earned in the competitive case to the monopoly firm.



Now, suppose that all the firms in the industry merge and a government restriction prohibits entry by any new firms. Our perfectly competitive industry is now a monopoly. Assume the monopoly continues to have the same marginal cost and demand curves that the competitive industry did. The monopoly firm faces the same market demand curve, from which it derives its marginal revenue curve. It maximizes profit at output Q_m and charges price P_m. Output is lower and price higher than in the competitive solution.

Society would gain by moving from the monopoly solution at Q_m to the competitive solution at Q_c. The benefit to consumers would be given by the area under the demand curve between Q_m and Q_c; it is the area Q_mRCQ_c. An increase in output, of course, has a cost. Because the marginal cost curve measures the cost of each additional unit, we can think of the area under the marginal cost curve over some range of output as measuring the total cost of that output. Thus, the total cost of increasing output from Q_m to Q_c is the area under the marginal cost curve over that range—the area Q_mGCQ_c. Subtracting this cost from the benefit gives us the net gain of moving from the monopoly to the competitive solution; it is the shaded area GRC. That is the potential gain from moving to the efficient solution. The area GRC is a deadweight loss.

Monopoly and Equity

The monopoly solution raises issues not just of efficiency but also of equity. Figure 10.11 shows that the monopolist charges price P_m rather than the competitive price P_c; the higher price charged by the monopoly firm reduces consumer surplus. Consumer surplus is the difference between what consumers are willing to pay for a good and what they actually pay. It is measured by the area under the demand curve and above the price of the good over the range of output produced.

If the industry were competitive, consumer surplus would be the area below the demand curve and above P_cC. With monopoly, consumer surplus would be the area below the demand curve and above P_mR. Part of the reduction in consumer surplus is the area under the demand curve between Q_c and

Q_m; it is contained in the deadweight loss area GRC. But consumers also lose the area of the rectangle bounded by the competitive and monopoly prices and by the monopoly output; this lost consumer surplus is transferred to the monopolist.

The fact that society suffers a deadweight loss due to monopoly is an efficiency problem. But the transfer of a portion of consumer surplus to the monopolist is an equity issue. Is such a transfer legitimate? After all, the monopoly firm enjoys a privileged position, protected by barriers to entry from competition. Should it be allowed to extract these gains from consumers? We will see that public policy suggests that the answer is no. Regulatory efforts imposed in monopoly cases often seek to reduce the degree to which monopoly firms extract consumer surplus from consumers by reducing the prices these firms charge.

Monopoly and the Concentration of Power

The objections to monopoly run much deeper than worries over economic efficiency and high prices. Because it enjoys barriers that block potential rivals, a monopoly firm wields considerable market power. For many people, that concentration of power is objectionable. A decentralized, competitive market constantly tests the ability of firms to satisfy consumers, pushes them to find new products and new and better production methods, and whittles away economic profits. Firms that operate in the shelter of monopoly may be largely immune to such pressures. Consumers are likely to be left with fewer choices, higher costs, and lower quality.

Perhaps more important in the view of many economists is the fact that the existence of economic profits provides both an incentive and the means for monopolists to aggressively protect their position and extend it if possible. These economists point out that monopolists may be willing to spend their economic profits in attempts to influence political leaders and public authorities (including regulatory authorities) who can help them maintain or enhance their monopoly position. Graft and corruption may be the result, claim these critics. Indeed, Microsoft has been accused by its rivals of bullying computer manufacturers into installing its web browser, Internet Explorer, exclusively on their computers.

Attitudes about Microsoft reflect these concerns. Even among people who feel that its products are good and fairly priced, there is uneasiness about our seeming dependence on them. And once it has secured its dominant position, will it charge more for its products? Will it continue to innovate?

Public Policy Toward Monopoly

Pulling together what we have learned in this chapter on monopoly and previously on perfect competition, Table 10.1 summarizes the differences between the models of perfect competition and monopoly. Most importantly we note that whereas the perfectly competitive firm is a price taker, the monopoly firm is a price setter. Because of this difference, we can object to monopoly on grounds of economic efficiency; monopolies produce too little and charge too much. Also, the high price and persistent profits strike many as inequitable. Others may simply see monopoly as an unacceptable concentration of power.

TABLE 10.1 Characteristics of Perfect Competition and Monopoly

Characteristic or Event	Perfect Competition	Monopoly
Market	Large number of sellers and buyers producing a homogeneous good or service, easy entry.	Large number of buyers, one seller. Entry is blocked.
Demand and marginal revenue curves	The firm's demand and marginal revenue curve is a horizontal line at the market price.	The firm faces the market demand curve; marginal revenue is below market demand.
Price	Determined by demand and supply; each firm is a price taker. Price equals marginal cost.	The monopoly firm determines price; it is a price setter. Price is greater than marginal cost.
Profit maximization	Firms produce where marginal cost equals marginal revenue	Firms produce where marginal cost equals marginal revenue and charge the corresponding price on the demand curve.
Profit	Entry forces economic profit to zero in the long run.	Because entry is blocked, a monopoly firm can sustain an economic profit in the long run.
Efficiency	The equilibrium solution is efficient because price equals marginal cost.	The equilibrium solution is inefficient because price is greater than marginal cost.

Public policy toward monopoly generally recognizes two important dimensions of the monopoly problem. On the one hand, the combining of competing firms into a monopoly creates an inefficient and, to many, inequitable solution. On the other hand, some industries are characterized as natural monopolies; production by a single firm allows economies of scale that result in lower costs.

The combining of competing firms into a monopoly firm or unfairly driving competitors out of business is generally forbidden in the United States. Regulatory efforts to prevent monopoly fall under the purview of the nation's antitrust laws, discussed in more detail in a later chapter.

At the same time, we must be careful to avoid the mistake of simply assuming that competition is the alternative to monopoly, that every monopoly can and should be replaced by a competitive market. One key source of monopoly power, after all, is economies of scale. In the case of natural monopoly, the alternative to a single firm is many small, high-cost producers. We may not like having only one local provider of water, but we might like even less having dozens of providers whose costs—and prices—are higher. Where monopolies exist because economies of scale prevail over the entire range of market demand, they may serve a useful economic role. We might want to regulate their production and pricing choices, but we may not want to give up their cost advantages.

Where a natural monopoly exists, the price charged by the firm and other aspects of its behavior may be subject to regulation. Water or natural gas, for example, are often distributed by a public utility—a monopoly firm—at prices regulated by a state or local government agency. Typically, such agencies seek to force the firm to charge lower prices, and to make less profit, than it would otherwise seek.

Although economists are hesitant to levy blanket condemnations of monopoly, they are generally sharply critical of monopoly power where no rationale for it exists. When firms have substantial monopoly power only as the result of government policies that block entry, there may be little defense for their monopoly positions.

Public policy toward monopoly aims generally to strike the balance implied by economic analysis. Where rationales exist, as in the case of natural monopoly, monopolies are permitted—and their prices are regulated. In other cases, monopoly is prohibited outright. Societies are likely to at least consider taking action of some kind against monopolies unless they appear to offer cost or other technological advantages.

3.2 The Fragility of Monopoly Power

An important factor in thinking about public policy toward monopoly is to recognize that monopoly power can be a fleeting thing. Firms constantly seek out the market power that monopoly offers. When conditions are right to achieve this power, firms that succeed in carving out monopoly positions enjoy substantial profits. But the potential for high profits invites continuing attempts to break down the barriers to entry that insulate monopolies from competition.

Technological change and the pursuit of profits chip away constantly at the entrenched power of monopolies. Breathtaking technological change has occurred in the telecommunications industry. Catalog companies are challenging the monopoly positions of some retailers; internet booksellers and online textbook companies such as Flatworldknowledge.com are challenging the monopoly power of your university's bookstore; and Federal Express, UPS, and other companies are taking on the U.S. Postal Service. The assaults on monopoly power are continuous. Thus, even the monopoly firm must be on the lookout for potential competitors.

Potential rivals are always beating at the door and thereby making the monopoly's fragile market contestable—that is, open to entry, at least in the sense of rival firms producing "close enough," if not perfect, substitutes—close enough that they might eliminate the firm's monopoly power.

KEY TAKEAWAYS

- A monopoly firm produces an output that is less than the efficient level. The result is a deadweight loss to society, given by the area between the demand and marginal cost curves over the range of output between the output chosen by the monopoly firm and the efficient output.
- The higher price charged by the monopoly firm compared to the perfectly competitive firm reduces consumer surplus, part of which is transferred to the monopolist. This transfer generates an equity issue.
- The monopoly firm's market power reduces consumers' choices and may result in higher prices, but there may be advantages to monopoly as well, such as economies of scale and technological innovations encouraged by the patent system.
- Public policy toward monopoly consists of antitrust laws and regulation of natural monopolies.
- Forces that limit the power of monopoly firms are the constant effort by other firms to capture some of the monopoly firm's profits and technological change that erodes monopoly power.

TRY IT!

Does the statement below better describe a firm operating in a perfectly competitive market or a firm that is a monopoly?

1. The demand curve faced by the firm is downward-sloping.
2. The demand curve and the marginal revenue curves are the same.
3. Entry and exit are relatively difficult.
4. The firm is likely to be concerned about antitrust laws.
5. Consumer surplus would be increased if the firm produced more output.

Case in Point: Technological Change, Public Policy, and Competition in Telecommunications

© 2010 Jupiterimages Corporation

Back in the olden days—before 1984—to use a telephone in the United States almost certainly meant being a customer of AT&T. Ma Bell, as the company was known, provided local and long-distance service to virtually every U.S. household. AT&T was clearly a monopoly.

The Justice Department began its battle with AT&T in the 1970s, charging it with monopolizing the industry. The case culminated in a landmark 1984 ruling that broke the company up into seven so-called "Baby Bells" that would provide local telephone service. AT&T would continue to provide long-distance service.

In effect, the ruling replaced a single national monopoly with seven regional monopolies in local telephone service. AT&T maintained its monopoly position in long-distance service—for a while. The turmoil that has followed illustrates the fragility of monopoly power.

Technological developments in the industry have brought dramatic changes. Companies like MCI found ways to challenge AT&T's monopoly position in long-distance telephone service. Cable operators sprang up, typically developing monopoly power over the provision of cable television in their regional markets. But, technological change has eroded the monopoly power of local telephone companies and of cable operators. Cable companies have begun providing telephone service in competition with local telephone companies, telephone companies have begun providing cable services as well as Internet access, and the introduction of wireless communications has further blurred the distinction between different types of companies. Now, the ready availability of video services on the Internet threatens to make cable providers outmoded middlemen.

The Diffusion Group, a firm that provides analysis of the telecommunications industry, predicts that by 2010 more than 300 million households worldwide will have high-speed Internet access and will thus be able to take advantage of increasing availability of video services on the Internet.

"When you go back to the early 1980s, we were talking about a single product," Duane Ackerman, CEO of BellSouth told *The Wall Street Journal*. "Since then, there has been a tremendous explosion of technology. It has changed everything."

Cable companies once had two-thirds of the market for television services. Today, telephone companies are making inroads into that market. In 2005, telephone companies picked up more television customers than cable companies did. The fierce competition has been reflected in the falling stock prices of cable companies. Comcast Corp., the largest cable company in the United States with 22 million subscribers, suffered a 22% reduction in its stock price in 2005. Mediacom, the seventh largest cable company, had a 33% reduction in its stock price between 2004 and 2006. By 2008, their stock prices have still not recovered.

Already, Time Warner, a cable company that charges its customers $39.95 per month, is offering customers that call to cancel their service to switch to another provider a discount to $29.95. "It's a save tactic," admits company spokesperson Mark Harrad. It may be that, but it surely appears to be a harbinger of what is coming to telecommunications.

Sources: Peter Grant and Amy, Schatz, "For Cable Giants, AT&T Deal Is One More Reason to Worry," The Wall Street Journal Online, March 7, 2006, A1. Dionne Searcey, Amy Schatz, Almar LaTour and Dennis K. Berman, "A Reborn AT&T to Buy BellSouth," The Wall Street Journal Online, March 6, 2006, A1. Sara Silver, Dennis K. Berman, and Leila cs Abroud, "In Lucent Deal, Two Rivals Face Inroads from China," The Wall Street Journal Online, March 25, 2006, A1;Sara Silver, Dennis K. Berman, and Leila cs Abroud, "In Lucent Deal, Two Rivals Face China," The Wall Street Journal Online, March 25, 2006, A1.

ANSWERS TO TRY IT! PROBLEMS

1. monopoly
2. perfect competition
3. monopoly
4. monopoly
5. monopoly

4. REVIEW AND PRACTICE

Summary

This chapter has examined the profit-maximizing behavior of monopoly firms. Monopoly occurs if an industry consists of a single firm and entry into that industry is blocked.

Potential sources of monopoly power include the existence of economies of scale over the range of market demand, locational advantages, high sunk costs associated with entry, restricted ownership of raw materials and inputs, and government restrictions such as licenses or patents. Network effects for certain products further increase the market power that patents afford.

Because the demand curve faced by the monopolist is downward-sloping, the firm is a price setter. It will maximize profits by producing the quantity of output at which marginal cost equals marginal revenue. The profit-maximizing price is then found on the demand curve for that quantity.

Because a typical monopolist holds market price above marginal cost, the major impact of monopoly is a reduction in efficiency. Compared to a competitive market, the monopoly is characterized by more centralized power, potential higher profits, and less pressure to be responsive to consumer preferences. Public policy toward monopoly includes antitrust laws and, in the case of natural monopolies, regulation of price and other aspects of the firm's behavior.

CONCEPT PROBLEMS

1. Consider the following firms. Would you regard any of them as a monopoly? Why or why not? Could you use the monopoly model in analyzing the choices of any of them? Explain.
    a. the best restaurant in town
    b. your barber or beautician
    c. your local telephone company
    d. your campus bookstore
    e. Microsoft
    f. Amtrak
    g. the United States Postal Service
2. Explain the difference between the demand curve facing a monopoly firm and the demand curve facing a perfectly competitive firm.
3. What are the necessary conditions for a monopoly position in the market to be established?
4. A monopoly firm is free to charge any price it wishes. What constrains its choice of a price?
5. Suppose the government were to impose an annual license fee on a monopolist that just happened to be equal to its economic profits for a particular year. How would such a fee affect price and output? Do you think that such a fee would be appropriate? Why or why not?
6. Name one monopoly firm you deal with. What is the source of its monopoly power? Do you think it seeks to maximize its profits?
7. "A monopolist will never produce so much output as to operate in the inelastic portion of the demand curve." Explain.
8. "A monopoly is not efficient, and its pricing behavior leads to losses to society." What does this statement mean? Should society ban monopolies?
9. A small town located 30 miles from the nearest town has only one service station. Is the service station a monopoly? Why or why not?
10. Explain why under monopoly price is greater than marginal revenue, while under perfect competition price is equal to marginal revenue.
11. In what sense can the monopoly equilibrium be considered inequitable?
12. What is a natural monopoly? Should a natural monopoly be allowed to exist?
13. Give some examples of industries in which you think natural monopoly conditions are likely to prevail. Why do you think so?
14. People often blame the high prices for events such as professional football and basketball and baseball games on the high salaries of professional athletes. Assuming one of these teams is a monopoly, use the model to refute this argument.
15. How do the following events affect a monopoly firm's price and output? How will it affect the firm's profits? Illustrate your answers graphically.
    a. an increase in labor costs in the market in which the firm operates
    b. a reduction in the price of gasoline
    c. the firm's Chief Executive Officer persuades the Board to increase his or her annual salary
    d. demand for the firm's product falls
    e. demand for the firm's product rises
    f. the price of a substitute for the firm's product rises

NUMERICAL PROBLEMS

1. A university football team estimates that it faces the demand schedule shown for tickets for each home game it plays. The team plays in a stadium that holds 60,000 fans. It estimates that its marginal cost of attendance, and thus for tickets sold, is zero.

Price per ticket	Tickets per game
$100	0
80	20,000
60	40,000
40	60,000
20	80,000
0	100,000

   a. Draw the demand and marginal revenue curves. Compute the team's profit-maximizing price and the number of tickets it will sell at that price.
   b. Determine the price elasticity of demand at the price you determined in part (a).
   c. How much total revenue will the team earn?
   d. Now suppose the city in which the university is located imposes a $10,000 annual license fee on all suppliers of sporting events, including the University. How does this affect the price of tickets?
   e. Suppose the team increases its spending for scholarships for its athletes. How will this affect ticket prices, assuming that it continues to maximize profit?
   f. Now suppose that the city imposes a tax of $10 per ticket sold. How would this affect the price charged by the team?

2. A monopoly firm faces a demand curve given by the following equation: $P = \$500 - 10Q$, where Q equals quantity sold per day. Its marginal cost curve is $MC = \$100$ per day. Assume that the firm faces no fixed cost.
   a. How much will the firm produce?
   b. How much will it charge?
   c. Can you determine its profit per day? (Hint: you can; state how much it is.)
   d. Suppose a tax of $1,000 per day is imposed on the firm. How will this affect its price?
   e. How would the $1,000 per day tax its output per day?
   f. How would the $1,000 per day tax affect its profit per day?
   g. Now suppose a tax of $100 per unit is imposed. How will this affect the firm's price?
   h. How would a $100 per unit tax affect the firm's profit maximizing output per day?
   i. How would the $100 per unit tax affect the firms profit per day?

CHAPTER 11

The World of Imperfect Competition

START UP: EBAY NEEDS GOOGLE, GOOGLE NEEDS EBAY, AND NEITHER TRUSTS THE OTHER

The Internet auction site eBay has had a close and cooperative relationship with Google, the giant search engine. eBay has relied heavily on Google to advertise its products. Google relies heavily on the advertising revenue it gets from eBay. The greater the success of eBay, the greater the revenue Google will have from eBay's advertising. The greater the success of Google as a search engine, the greater will be the impact of eBay's advertising. To paraphrase Rick's line from Casablanca, "This could be a beautiful relationship." It is not. The two Internet giants simply do not get along.

Consider what happened in 2007. A Google spokesman said the firm was hosting a "Freedom Party" to announce the inauguration of a new payments service that would compete directly with PayPal, the online payment service owned by eBay. eBay was quick to retaliate. It pulled all of its advertising from Google later on the same day Google made its announcement. Two days later, Google backed down. It canceled its party and the payment service the party was to kick off.[1]

In 2003, eBay had commissioned an analysis of whether Google represented a threat to its operations. The study concluded that Google was unlikely to enter into e-commerce and was not a potential rival to eBay. That sanguine conclusion started to unravel in 2005. Google began recruiting eBay engineers. In October, Google started testing Google Base, a free classified advertising service that threatened eBay's auction service.

Executives at eBay took the threat seriously. In private meetings, they divided into two teams. A green team represented eBay's interests; a red team tried to emulate Google's strategy. The red team concluded that Google represented a serious threat, and eBay executives began exploratory talks with Microsoft and Yahoo to see if some collaborative effort could ward off the Google threat.

eBay spokesman Chris Donlay describes the firm's dilemma of dealing with a firm that has been a valuable ally but at the same time could be a competitive threat. "Given how really fast the Internet changes, it comes as no surprise that the line between competition and cooperation is sometimes blurry."

By the late spring of 2006, eBay's management was still in a quandary about what to do about Google. Some executives, fearful of losing the advantages of continuing to work with Google, want to maintain eBay's ties to the firm. Others worried that continuing a close relationship with Google was akin to putting the fox in the proverbial henhouse. They want to move quickly to establish a relationship with Yahoo or with Microsoft that would compete with Google.[2]

The tension between eBay and Google hardly suggests the aloof world of perfect competition where consumers are indifferent about which firm has produced a particular product, where each firm knows it can sell all it wants at the going market price, where firms must settle for zero economic profit in the long run. Nor is it the world of monopoly, where a single firm maximizes its profits, believing that barriers to entry will keep out would-be competitors, at least for a while. This is the world of imperfect competition, one that lies between the idealized extremes of perfect competition and monopoly. It is a world in which firms battle over market shares, in which economic profits may persist, in which rivals try to outguess each other with pricing, advertising, and product-development strategies.

Unlike the chapters on perfect competition and monopoly, this chapter does not provide a single model to explain firms' behavior. There are too many variations on an uncertain theme for one model to explain the complexities of imperfect competition. Rather, the chapter provides an overview of some of the many different models and explanations advanced by economists for the behavior of firms in the imperfectly competitive markets. The analytical tools you have acquired in the course of studying the models of competitive and monopoly markets will be very much in evidence in this discussion.

imperfect competition

A market structure with more than one firm in an industry in which at least one firm is a price setter.

The spectrum of business enterprise ranges from perfectly competitive firms to monopoly. Between these extremes lies the business landscape in which the vast majority of firms—those in the world of imperfect competition—actually operate. **Imperfect competition** is a market structure with more than one firm in an industry in which at least one firm is a price setter. An imperfectly competitive firm has a degree of monopoly power, either based on product differentiation that leads to a downward-sloping demand curve or resulting from the interaction of rival firms in an industry with only a few firms.

There are two broad categories of imperfectly competitive markets. The first is one in which many firms compete, each offering a slightly different product. The second is one in which the industry is dominated by a few firms. Important features of both kinds of markets are advertising and price discrimination, which we examine later in this chapter.

1. MONOPOLISTIC COMPETITION: COMPETITION AMONG MANY

LEARNING OBJECTIVES

1. **Explain the main characteristics of a monopolistically competitive industry, describing both its similarities and differences from the models of perfect competition and monopoly.**
2. **Explain and illustrate both short-run equilibrium and long-run equilibrium for a monopolistically competitive firm.**
3. **Explain what it means to say that a firm operating under monopolistic competition has excess capacity in the long run and discuss the implications of this conclusion.**

monopolistic competition

A model characterized by many firms producing similar but differentiated products in a market with easy entry and exit.

The first model of an imperfectly competitive industry that we shall investigate has conditions quite similar to those of perfect competition. The model of monopolistic competition assumes a large number of firms. It also assumes easy entry and exit. This model differs from the model of perfect competition in one key respect: it assumes that the goods and services produced by firms are differentiated. This differentiation may occur by virtue of advertising, convenience of location, product quality, reputation of the seller, or other factors. Product differentiation gives firms producing a particular product some degree of price-setting or monopoly power. However, because of the availability of close substitutes, the price-setting power of monopolistically competitive firms is quite limited. **Monopolistic competition** is a model characterized by many firms producing similar but differentiated products in a market with easy entry and exit.

Restaurants are a monopolistically competitive sector; in most areas there are many firms, each is different, and entry and exit are very easy. Each restaurant has many close substitutes—these may include other restaurants, fast-food outlets, and the deli and frozen-food sections at local supermarkets. Other industries that engage in monopolistic competition include retail stores, barber and beauty shops, auto-repair shops, service stations, banks, and law and accounting firms.

1.1 Profit Maximization

Suppose a restaurant raises its prices slightly above those of similar restaurants with which it competes. Will it have any customers? Probably. Because the restaurant is different from other restaurants, some people will continue to patronize it. Within limits, then, the restaurant can set its own prices; it does not take the market prices as given. In fact, differentiated markets imply that the notion of a single "market price" is meaningless.

Because products in a monopolistically competitive industry are differentiated, firms face downward-sloping demand curves. Whenever a firm faces a downward-sloping demand curve, the graphical framework for monopoly can be used. In the short run, the model of monopolistic competition looks exactly like the model of monopoly. An important distinction between monopoly and monopolistic competition, however, emerges from the assumption of easy entry and exit. In monopolistic competition, entry will eliminate any economic profits in the long run. We begin with an analysis of the short run.

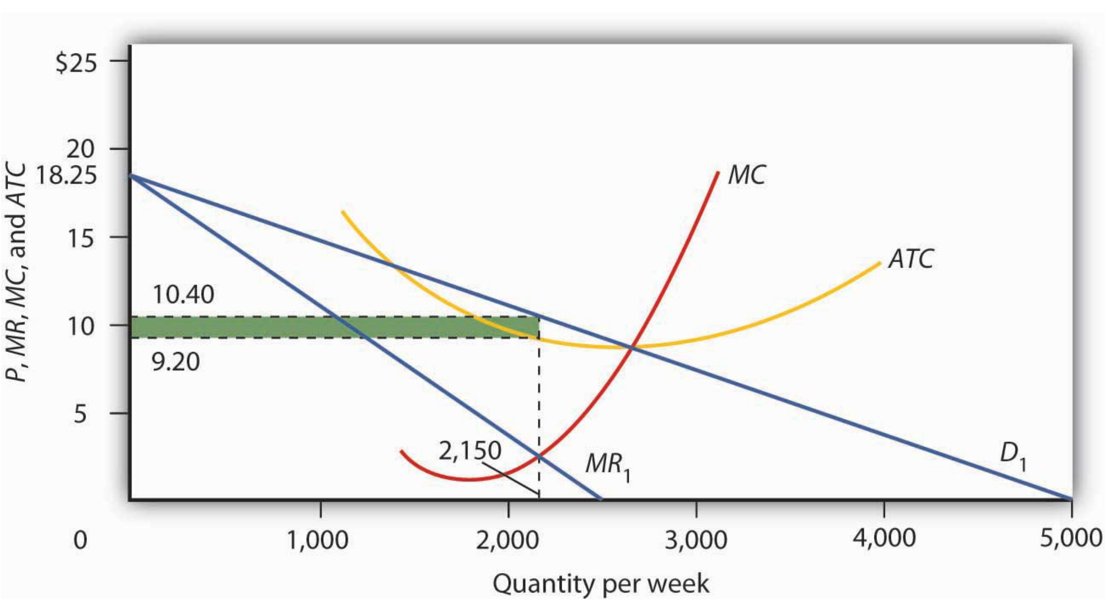
The Short Run

Because a monopolistically competitive firm faces a downward-sloping demand curve, its marginal revenue curve is a downward-sloping line that lies below the demand curve, as in the monopoly model. We can thus use the model of monopoly that we have already developed to analyze the choices of a monopsony in the short run.

Figure 11.1 shows the demand, marginal revenue, marginal cost, and average total cost curves facing a monopolistically competitive firm, Mama's Pizza. Mama's competes with several other similar firms in a market in which entry and exit are relatively easy. Mama's demand curve D_1 is downward-sloping; even if Mama's raises its prices above those of its competitors, it will still have some customers. Given the downward-sloping demand curve, Mama's marginal revenue curve MR_1 lies below demand. To sell more pizzas, Mama's must lower its price, and that means its marginal revenue from additional pizzas will be less than price.

FIGURE 11.1 Short-Run Equilibrium in Monopolistic Competition

Looking at the intersection of the marginal revenue curve MR_1 and the marginal cost curve MC, we see that the profit-maximizing quantity is 2,150 units per week. Reading up to the average total cost curve ATC, we see that the cost per unit equals \$9.20. Price, given on the demand curve D_1, is \$10.40, so the profit per unit is \$1.20. Total profit per week equals \$1.20 times 2,150, or \$2,580; it is shown by the shaded rectangle.



Given the marginal revenue curve *MR* and marginal cost curve *MC,* Mama's will maximize profits by selling 2,150 pizzas per week. Mama's demand curve tells us that it can sell that quantity at a price of $10.40. Looking at the average total cost curve *ATC,* we see that the firm's cost per unit is $9.20. Its economic profit per unit is thus $1.20. Total economic profit, shown by the shaded rectangle, is $2,580 per week.

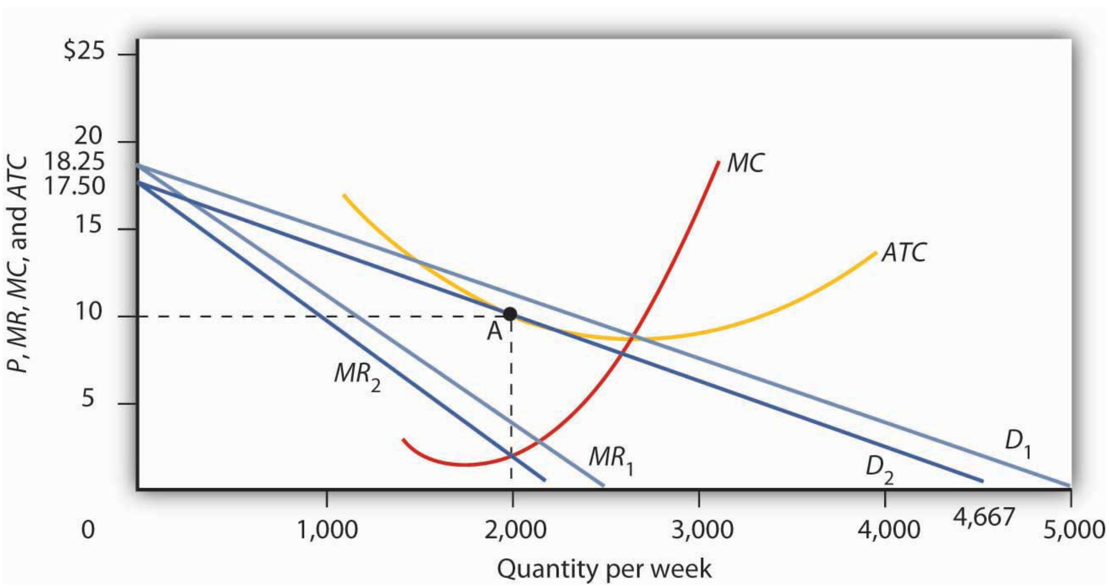
The Long Run

We see in Figure 11.1 that Mama's Pizza is earning an economic profit. If Mama's experience is typical, then other firms in the market are also earning returns that exceed what their owners could be earning in some related activity. Positive economic profits will encourage new firms to enter Mama's market.

As new firms enter, the availability of substitutes for Mama's pizzas will increase, which will reduce the demand facing Mama's Pizza and make the demand curve for Mama's Pizza more elastic. Its demand curve will shift to the left. Any shift in a demand curve shifts the marginal revenue curve as well. New firms will continue to enter, shifting the demand curves for existing firms to the left, until pizza firms such as Mama's no longer make an economic profit. The zero-profit solution occurs where Mama's demand curve is tangent to its average total cost curve—at point A in Figure 11.2. Mama's price will fall to $10 per pizza and its output will fall to 2,000 pizzas per week. Mama's will just cover its opportunity costs, and thus earn zero economic profit. At any other price, the firm's cost per unit would be greater than the price at which a pizza could be sold, and the firm would sustain an economic loss. Thus, the firm and the industry are in long-run equilibrium. There is no incentive for firms to either enter or leave the industry.

FIGURE 11.2 Monopolistic Competition in the Long Run

The existence of economic profits in a monopolistically competitive industry will induce entry in the long run. As new firms enter, the demand curve D_1 and marginal revenue curve MR_1 facing a typical firm will shift to the left, to D_2 and MR_2. Eventually, this shift produces a profit-maximizing solution at zero economic profit, where D_2 is tangent to the average total cost curve *ATC* (point A). The long-run equilibrium solution here is an output of 2,000 units per week at a price of $10 per unit.



Had Mama's Pizza and other similar restaurants been incurring economic losses, the process of moving to long-run equilibrium would work in reverse. Some firms would exit. With fewer substitutes available, the demand curve faced by each remaining firm would shift to the right. Price and output at each restaurant would rise. Exit would continue until the industry was in long-run equilibrium, with the typical firm earning zero economic profit.

Such comings and goings are typical of monopolistic competition. Because entry and exit are easy, favorable economic conditions in the industry encourage start-ups. New firms hope that they can differentiate their products enough to make a go of it. Some will; others will not. Competitors to Mama's may try to improve the ambience, play different music, offer pizzas of different sizes and types. It might take a while for other restaurants to come up with just the right product to pull customers and profits away from Mama's. But as long as Mama's continues to earn economic profits, there will be incentives for other firms to try.

Heads Up!

The term "monopolistic competition" is easy to confuse with the term "monopoly." Remember, however, that the two models are characterized by quite different market conditions. A monopoly is a single firm with high barriers to entry. Monopolistic competition implies an industry with many firms, differentiated products, and easy entry and exit.

Why is the term monopolistic competition used to describe this type of market structure? The reason is that it bears some similarities to both perfect competition and to monopoly. Monopolistic competition is similar to perfect competition in that in both of these market structures many firms make up the industry and entry and exit are fairly easy. Monopolistic competition is similar to monopoly in that, like monopoly firms, monopolistically competitive firms have at least some discretion when it comes to setting prices. However, because monopolistically competitive firms produce goods that are close substitutes for those of rival firms, the degree of monopoly power that monopolistically competitive firms possess is very low.

1.2 Excess Capacity: The Price of Variety

The long-run equilibrium solution in monopolistic competition always produces zero economic profit at a point to the left of the minimum of the average total cost curve. That is because the zero profit solution occurs at the point where the downward-sloping demand curve is tangent to the average total cost curve, and thus the average total cost curve is itself downward-sloping. By expanding output, the firm could lower average total cost. The firm thus produces less than the output at which it would minimize average total cost. A firm that operates to the left of the lowest point on its average total cost curve has **excess capacity**.

excess capacity

Situation in which a firm operates to the left of the lowest point on its average total cost curve.

Because monopolistically competitive firms charge prices that exceed marginal cost, monopolistic competition is inefficient. The marginal benefit consumers receive from an additional unit of the good is given by its price. Since the benefit of an additional unit of output is greater than the marginal cost, consumers would be better off if output were expanded. Furthermore, an expansion of output would reduce average total cost. But monopolistically competitive firms will not voluntarily increase output, since for them, the marginal revenue would be less than the marginal cost.

One can thus criticize a monopolistically competitive industry for falling short of the efficiency standards of perfect competition. But monopolistic competition is inefficient because of product differentiation. Think about a monopolistically competitive activity in your area. Would consumers be better off if all the firms in this industry produced identical products so that they could match the assumptions of perfect competition? If identical products were impossible, would consumers be better off if some of the firms were ordered to shut down on grounds the model predicts there will be "too many" firms? The inefficiency of monopolistic competition may be a small price to pay for a wide range of product choices. Furthermore, remember that perfect competition is merely a model. It is not a goal toward which an economy might strive as an alternative to monopolistic competition.

KEY TAKEAWAYS

- A monopolistically competitive industry features some of the same characteristics as perfect competition: a large number of firms and easy entry and exit.
- The characteristic that distinguishes monopolistic competition from perfect competition is differentiated products; each firm is a price setter and thus faces a downward-sloping demand curve.
- Short-run equilibrium for a monopolistically competitive firm is identical to that of a monopoly firm. The firm produces an output at which marginal revenue equals marginal cost and sets its price according to its demand curve.
- In the long run in monopolistic competition any economic profits or losses will be eliminated by entry or by exit, leaving firms with zero economic profit.
- A monopolistically competitive industry will have some excess capacity; this may be viewed as the cost of the product diversity that this market structure produces.

TRY IT!

Suppose the monopolistically competitive restaurant industry in your town is in long-run equilibrium, when difficulties in hiring cause restaurants to offer higher wages to cooks, servers, and dishwashers. Using graphs similar to Figure 11.1 and Figure 11.2, explain the effect of the wage increase on the industry in the short run and in the long run. Be sure to include in your answer an explanation of what happens to price, output, and economic profit.

Case in Point: Craft Brewers: The Rebirth of a Monopolistically Competitive Industry

© 2010 Jupiterimages Corporation

In the early 1900s, there were about 2,000 local beer breweries across America. Prohibition in the 1920s squashed the industry; after the repeal of Prohibition, economies of scale eliminated smaller breweries. By the early 1980s only about 40 remained in existence.

But the American desire for more variety has led to the rebirth of the nearly defunct industry. To be sure, large, national beer companies dominated the overall ale market in 1980 and they still do today, with 43 large national and regional breweries sharing about 85% of the U.S. market for beer. But their emphasis on similarly tasting, light lagers (at least, until they felt threatened enough by the new craft brewers to come up with their own specialty brands) left many niches to be filled. One niche was filled by imports, accounting for about 12% of the U.S. market. That leaves 3 to 4% of the national market for the domestic specialty or "craft" brewers.

The new craft brewers, which include contract brewers, regional specialty brewers, microbreweries, and brewpubs, offer choice. As Neal Leathers at Big Sky Brewing Company in Missoula, Montana put it, "We sort of convert people. If you haven't had very many choices, and all of a sudden you get choices—especially if those choices involve a lot of flavor and quality—it's hard to go back."

Aided by the recent legalization in most states of brewpubs, establishments where beers are manufactured and retailed on the same premises, the number of microbreweries grew substantially over the last 25 years. A recent telephone book in Colorado Springs, a city with a population of about a half million and the home of the authors of your textbook, listed nine microbreweries and brewpubs; more exist, but prefer to be listed as restaurants.

To what extent does this industry conform to the model of monopolistic competition? Clearly, the microbreweries sell differentiated products, giving them some degree of price-setting power. A sample of four brewpubs in the downtown area of Colorado Springs revealed that the price of a house beer ranged from 13 to 22 cents per ounce.

Entry into the industry seems fairly easy, judging from the phenomenal growth of the industry. After more than a decade of explosive growth and then a period of leveling off, the number of craft breweries, as they are referred to by the Association of Brewers, stood at 1,463 in 2007. The start-up cost ranges from $100,000 to $400,000, according to Kevin Head, the owner of the Rhino Bar, also in Missoula.

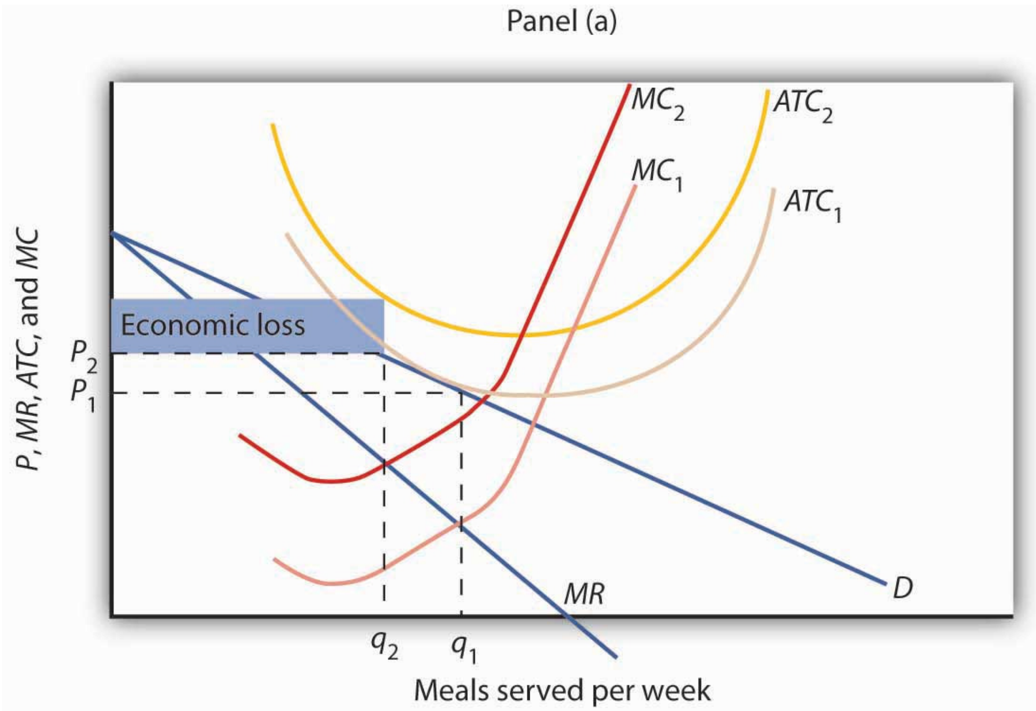
The monopolistically competitive model also predicts that while firms can earn positive economic profits in the short run, entry of new firms will shift the demand curve facing each firm to the left and economic profits will fall toward zero. Some firms will exit as competitors win customers away from them. In the combined microbrewery and brewpub sub-sectors of the craft beer industry in 2007, for example, there were 94 openings and 51 closings.

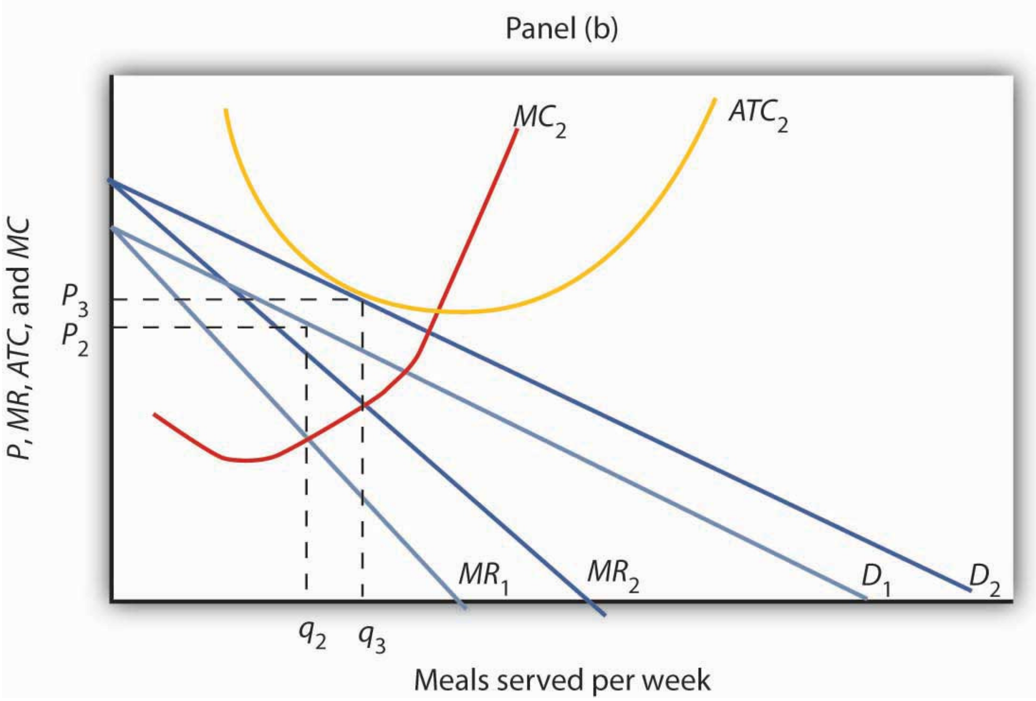
Sources: Jim Ludwick, "The Art of Zymurgy—It's the Latest Thing: Microbrewers are Tapping into the New Specialty Beer Market," Missoulian (November 29, 1996): p. A1; Brewers Association, "2007 Craft Beer Industry Statistics," April 17, 2008, www.beertown.org/craftbrewing/statistics.html.

ANSWER TO TRY IT! PROBLEM

As shown in Panel (a), higher wages would cause both *MC* and *ATC* to increase. The upward shift in *MC* from MC_1 to MC_2 would cause the profit-maximizing level of output (number of meals served per week, in this case) to fall from q_1 to q_2 and price to increase from P_1 to P_2. The increase in *ATC* from ATC_1 to ATC_2 would mean that some restaurants would be earning negative economic profits, as shown by the shaded area.

As shown in Panel (b), in the long run, as some restaurants close down, the demand curve faced by the typical remaining restaurant would shift to the right from *D* to *D′*. The demand curve shift leads to a corresponding shift in marginal revenue from *MR* to MR_1. Price would increase further from P_2 to P_3, and output would increase to q_3, above q_2. In the new long-run equilibrium, restaurants would again be earning zero economic profit.





2. OLIGOPOLY: COMPETITION AMONG THE FEW

LEARNING OBJECTIVES

1. **Explain the main characteristics of an oligopoly, differentiating it from other types of market structures.**
2. **Explain the measures that are used to determine the degree of concentration in an industry.**
3. **Explain and illustrate the collusion model of oligopoly.**
4. **Discuss how game theory can be used to understand the behavior of firms in an oligopoly.**

In July, 2005, General Motors Corporation (GMC) offered "employee discount pricing" to virtually all GMC customers, not just employees and their relatives. This new marketing strategy introduced by GMC obviously affected Ford, Chrysler, Toyota and other automobile and truck manufacturers; Ford matched GMC's employee-discount plan by offering up to $1,000 to its own employees who convinced friends to purchase its cars and trucks. Ford also offered its customers the same prices paid by its employees. By mid-July, Chrysler indicated that it was looking at many alternatives, but was waiting for GMC to make its next move. Ultimately, Chrysler also offered employee discount pricing.

Toyota had to respond. It quickly developed a new marketing strategy of its own, which included lowering the prices of its cars and offering new financing terms. The responses of Ford, Chrysler, and Toyota to GMC's pricing strategy obviously affected the outcome of that strategy. Similarly, a decision by Procter & Gamble to lower the price of Crest toothpaste may elicit a response from Colgate-Palmolive, and that response will affect the sales of Crest. In an **oligopoly**, the fourth and final market structure that we will study, the market is dominated by a few firms, each of which recognizes that its own actions will produce a response from its rivals and that those responses will affect it.

oligopoly

Situation in which a market is dominated by a few firms, each of which recognizes that its own actions will produce a response from its rivals and that those responses will affect it.

The firms that dominate an oligopoly recognize that they are interdependent: What one firm does affects each of the others. This interdependence stands in sharp contrast to the models of perfect competition and monopolistic competition, where we assume that each firm is so small that it assumes the rest of the market will, in effect, ignore what it does. A perfectly competitive firm responds to the market, not to the actions of any other firm. A monopolistically competitive firm responds to its own demand, not to the actions of specific rivals. These presumptions greatly simplify the analysis of perfect competition and monopolistic competition. We do not have that luxury in oligopoly, where the interdependence of firms is the defining characteristic of the market.

Some oligopoly industries make standardized products: steel, aluminum, wire, and industrial tools. Others make differentiated products: cigarettes, automobiles, computers, ready-to-eat breakfast cereal, and soft drinks.

2.1 Measuring Concentration in Oligopoly

Oligopoly means that a few firms dominate an industry. But how many is "a few," and how large a share of industry output does it take to "dominate" the industry?

Compare, for example, the ready-to-eat breakfast cereal industry and the ice cream industry. The cereal market is dominated by two firms, Kellogg's and General Mills, which together hold more than half the cereal market. This oligopoly operates in a highly concentrated market. The market for ice cream, where the four largest firms account for just less than a third of output, is much less concentrated.

One way to measure the degree to which output in an industry is concentrated among a few firms is to use a **concentration ratio**, which reports the percentage of output accounted for by the largest firms in an industry. The higher the concentration ratio, the more the firms in the industry take account of their rivals' behavior. The lower the concentration ratio, the more the industry reflects the characteristics of monopolistic competition or perfect competition.

concentration ratio

The percentage of output accounted for by the largest firms in an industry.

The U.S. Census Bureau, based on surveys it conducts of manufacturing firms every five years, reports concentration ratios. These surveys show concentration ratios for the largest 4, 8, 20, and 50 firms in each industry category. Some concentration ratios from the 2002 survey, the latest available, are reported in Table 11.1. Notice that the four-firm concentration ratio for breakfast cereals is 78%; for ice cream it is 48%.

TABLE 11.1 Concentration Ratios and Herfindahl–Hirschman Indexes

Two measures of industry concentration are reported by the Census Bureau: concentration ratios and the Herfindahl–Hirschman Index (HHI).

Industry	Largest 4 firms	Largest 8 firms	Largest 20 firms	Largest 50 firms	HHI
Ice cream	48	64	82	93	736
Breakfast cereals	78	91	99	100	2521
Cigarettes	95	99	100		*D
Men's and boys' shirts	38	53	73	89	481
Women's and girls' blouses and shirts	21	32	49	70	186
Automobiles	76	94	99	100	1911
Sporting and athletic goods	23	32	46	62	182
Dental laboratories	13	18	23	30	54
***D, data withheld by the government to avoid revealing information about specific firms.**					

Source: Selected statistics from Sector 31: Manufacturing: Subject Series—Concentration Ratios: Share of Value of Shipments Accounted for by the 4, 8, 20, and 50 Largest Companies for Industries: 2002 at http://www.census.gov/epcd/www/concentration.html.

Herfindahl–Hirschman Index

An alternative measure of concentration found by squaring the percentage share (stated as a whole number) of each firm in an industry, then summing these squared market shares.

An alternative measure of concentration is found by squaring the percentage share (stated as a whole number) of each firm in an industry, then summing these squared market shares to derive a **Herfindahl–Hirschman Index (HHI)**. The largest HHI possible is the case of monopoly, where one firm has 100% of the market; the index is 100^2, or 10,000. An industry with two firms, each with 50% of total output, has an HHI of 5,000 ($50^2 + 50^2$). In an industry with 10,000 firms that have 0.01% of the market each, the HHI is 1. Herfindahl–Hirschman Indexes reported by the Census Bureau are also given in Table 11.1. Notice that the HHI is 2,521 for breakfast cereals and only 736 for ice cream, suggesting that the ice cream industry is more competitive than the breakfast cereal industry.

In some cases, the census data understate the degree to which a few firms dominate the market. One problem is that industry categories may be too broad to capture significant cases of industry dominance. The sporting goods industry, for example, appears to be highly competitive if we look just at measures of concentration, but markets for individual goods, such as golf clubs, running shoes, and tennis rackets, tend to be dominated by a few firms. Further, the data reflect shares of the national market. A tendency for regional domination does not show up. For example, the concrete industry appears to be highly competitive. But concrete is produced in local markets—it is too expensive to ship it very far—and many of these local markets are dominated by a handful of firms.

The census data can also overstate the degree of actual concentration. The "automobiles" category, for example, has a four-firm concentration ratio that suggests the industry is strongly dominated by four large firms (in fact, U.S. production is dominated by three: General Motors, Ford, and Chrysler). Those firms hardly account for all car sales in the United States, however, as other foreign producers have captured a large portion of the domestic market. Including those foreign competitors suggests a far less concentrated industry than the census data imply.

2.2 The Collusion Model

There is no single model of profit-maximizing oligopoly behavior that corresponds to economists' models of perfect competition, monopoly, and monopolistic competition. Uncertainty about the interaction of rival firms makes specification of a single model of oligopoly impossible. Instead, economists have devised a variety of models that deal with the uncertain nature of rivals' responses in different ways. In this section we review one type of oligopoly model, the collusion model. After examining this traditional approach to the analysis of oligopoly behavior, we shall turn to another method of examining oligopolistic interaction: game theory.

Firms in any industry could achieve the maximum profit attainable if they all agreed to select the monopoly price and output and to share the profits. One approach to the analysis of oligopoly is to assume that firms in the industry collude, selecting the monopoly solution.

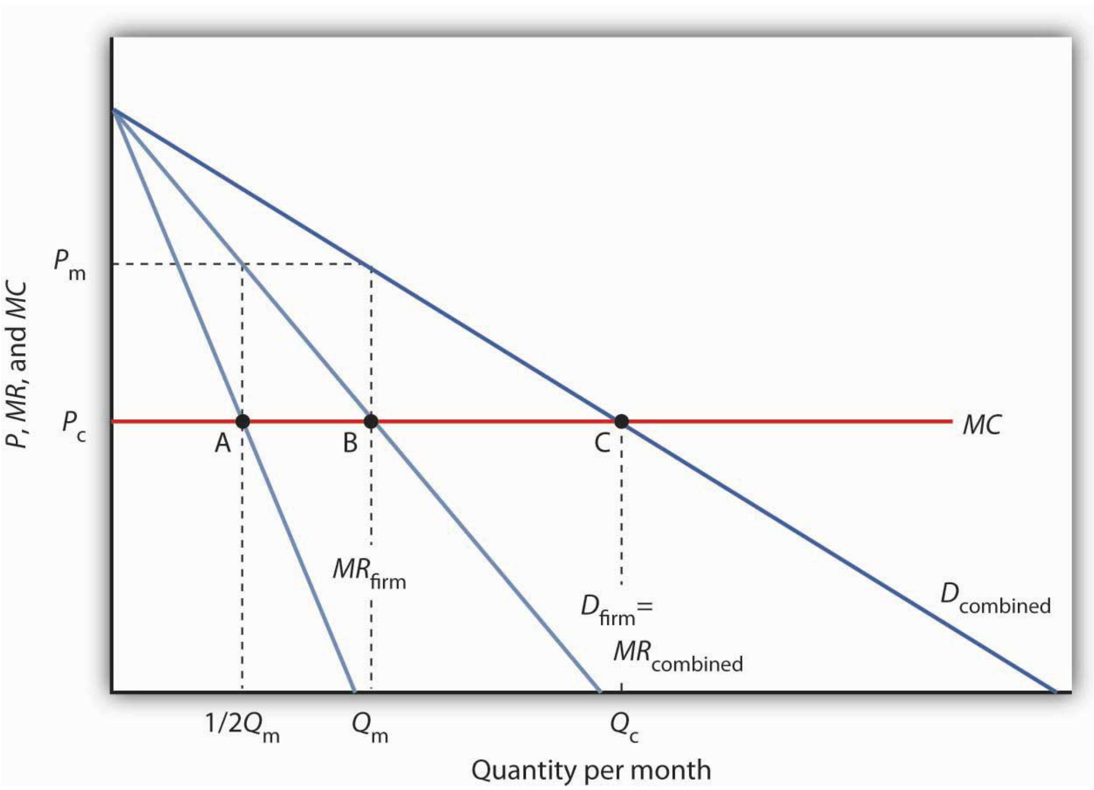
duopoly

An industry that has only two firms.

Suppose an industry is a **duopoly**, an industry with two firms. Figure 11.5 shows a case in which the two firms are identical. They sell identical products and face identical demand and cost conditions. To simplify the analysis, we will assume that each has a horizontal marginal cost curve, *MC*. The demand and marginal revenue curves are the same for both firms. We find the combined demand curve for the two firms, D_{combined}, by adding the individual demand curves together. Because one firm's demand curve, D_{firm}, represents one-half of market demand, it is the same as the combined marginal revenue curve for the two firms. If these two firms act as a monopoly, together they produce Q_m and charge a price P_m. This result is achieved if each firm selects its profit-maximizing output, which equals 1/2 Q_m. This solution is inefficient; the efficient solution is price P_c and output Q_c, found where the combined market demand curve $D_{combined}$ and the marginal cost curve MC intersect.

FIGURE 11.5 Monopoly Through Collusion

Two identical firms have the same horizontal marginal cost curve MC. Their demand curves D_{firm} and marginal revenue curves MR_{firm} are also identical. The combined demand curve is $D_{combined}$; the combined marginal revenue curve is $MR_{combined}$. The profits of the two firms are maximized if each produces 1/2 Q_m at point A. Industry output at point B is thus Q_m and the price is P_m. At point C, the efficient solution output would be Q_c, and the price would equal MC.



In the simplest form of collusion, **overt collusion**, firms openly agree on price, output, and other decisions aimed at achieving monopoly profits. Firms that coordinate their activities through overt collusion and by forming collusive coordinating mechanisms make up a **cartel**.

Firms form a cartel to gain monopoly power. A successful cartel can earn large profits, but there are several problems with forming and maintaining one. First, in many countries, including the United States, cartels are generally illegal.[3] They are banned, because their purpose is to raise prices and restrict output. Second, the cartel may not succeed in inducing all firms in the industry to join. Firms that remain outside the cartel can compete by lowering price, and thus they prevent the cartel from achieving the monopoly solution. Third, there is always an incentive for individual members to cheat on cartel agreements. Suppose the members of a cartel have agreed to impose the monopoly price in their market and to limit their output accordingly. Any one firm might calculate that it could charge slightly less than the cartel price and thus capture a larger share of the market for itself. Cheating firms expand output and drive prices down below the level originally chosen.

The Organization of Petroleum Exporting Countries (OPEC), perhaps the best-known cartel, is made up of 13 oil-producing countries. In the 1970s, OPEC successfully acted like a monopoly by restricting output and raising prices. By the mid-1980s, however, the monopoly power of the cartel had been weakened by expansion of output by nonmember producers such as Mexico and Norway and by cheating among the cartel members.

overt collusion

When firms openly agree on price, output, and other decisions aimed at achieving monopoly profits.

cartel

Firms that coordinate their activities through overt collusion and by forming collusive coordinating mechanisms.

tacit collusion

An unwritten, unspoken understanding through which firms agree to limit their competition.

An alternative to overt collusion is **tacit collusion**, an unwritten, unspoken understanding through which firms agree to limit their competition. Firms may, for example, begin following the price leadership of a particular firm, raising or lowering their prices when the leader makes such a change. The price leader may be the largest firm in the industry, or it may be a firm that has been particularly good at assessing changes in demand or cost. At various times, tacit collusion has been alleged to occur in a wide range of industries, including steel, cars, and breakfast cereals.

It is difficult to know how common tacit collusion is. The fact that one firm changes its price shortly after another one does cannot prove that a tacit conspiracy exists. After all, we expect to see the prices of all firms in a perfectly competitive industry moving together in response to changes in demand or production costs.

2.3 Game Theory and Oligopoly Behavior

strategic choice

A choice based on the recognition that the actions of others will affect the outcome of the choice and that takes these possible actions into account.

game theory

An analytical approach through which strategic choices can be assessed.

payoff

The outcome of a strategic decision.

Oligopoly presents a problem in which decision makers must select strategies by taking into account the responses of their rivals, which they cannot know for sure in advance. The Start Up feature at the beginning of this chapter suggested the uncertainty eBay faces as it considers the possibility of competition from Google. A choice based on the recognition that the actions of others will affect the outcome of the choice and that takes these possible actions into account is called a **strategic choice**. **Game theory** is an analytical approach through which strategic choices can be assessed.

Among the strategic choices available to an oligopoly firm are pricing choices, marketing strategies, and product-development efforts. An airline's decision to raise or lower its fares—or to leave them unchanged—is a strategic choice. The other airlines' decision to match or ignore their rival's price decision is also a strategic choice. IBM boosted its share in the highly competitive personal computer market in large part because a strategic product-development strategy accelerated the firm's introduction of new products.

Once a firm implements a strategic decision, there will be an outcome. The outcome of a strategic decision is called a **payoff**. In general, the payoff in an oligopoly game is the change in economic profit to each firm. The firm's payoff depends partly on the strategic choice it makes and partly on the strategic choices of its rivals. Some firms in the airline industry, for example, raised their fares in 2005, expecting to enjoy increased profits as a result. They changed their strategic choices when other airlines chose to slash their fares, and all firms ended up with a payoff of lower profits—many went into bankruptcy.

We shall use two applications to examine the basic concepts of game theory. The first examines a classic game theory problem called the prisoners' dilemma. The second deals with strategic choices by two firms in a duopoly.

The Prisoners' Dilemma

Suppose a local district attorney (DA) is certain that two individuals, Frankie and Johnny, have committed a burglary, but she has no evidence that would be admissible in court.

The DA arrests the two. On being searched, each is discovered to have a small amount of cocaine. The DA now has a sure conviction on a possession of cocaine charge, but she will get a conviction on the burglary charge only if at least one of the prisoners confesses and implicates the other.

The DA decides on a strategy designed to elicit confessions. She separates the two prisoners and then offers each the following deal: "If you confess and your partner doesn't, you will get the minimum sentence of one year in jail on the possession and burglary charges. If you both confess, your sentence will be three years in jail. If your partner confesses and you do not, the plea bargain is off and you will get six years in prison. If neither of you confesses, you will each get two years in prison on the drug charge."

The two prisoners each face a dilemma; they can choose to confess or not confess. Because the prisoners are separated, they cannot plot a joint strategy. Each must make a strategic choice in isolation.

The outcomes of these strategic choices, as outlined by the DA, depend on the strategic choice made by the other prisoner. The payoff matrix for this game is given in Figure 11.6. The two rows represent Frankie's strategic choices; she may confess or not confess. The two columns represent Johnny's strategic choices; he may confess or not confess. There are four possible outcomes: Frankie and Johnny both confess (cell A), Frankie confesses but Johnny does not (cell B), Frankie does not confess but Johnny does (cell C), and neither Frankie nor Johnny confesses (cell D). The portion at the lower left in each cell shows Frankie's payoff; the shaded portion at the upper right shows Johnny's payoff.

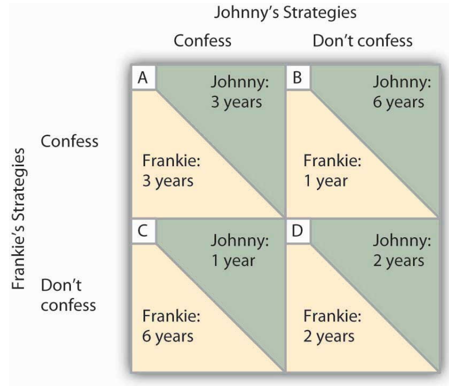
If Johnny confesses, Frankie's best choice is to confess—she will get a three-year sentence rather than the six-year sentence she would get if she did not confess. If Johnny does not confess, Frankie's best strategy is still to confess—she will get a one-year rather than a two-year sentence. In this game, Frankie's best strategy is to confess, regardless of what Johnny does. When a player's best strategy is the same regardless of the action of the other player, that strategy is said to be a **dominant strategy**. Frankie's dominant strategy is to confess to the burglary.

For Johnny, the best strategy to follow, if Frankie confesses, is to confess. The best strategy to follow if Frankie does not confess is also to confess. Confessing is a dominant strategy for Johnny as well. A game in which there is a dominant strategy for each player is called a **dominant strategy equilibrium**. Here, the dominant strategy equilibrium is for both prisoners to confess; the payoff will be given by cell A in the payoff matrix.

From the point of view of the two prisoners together, a payoff in cell D would have been preferable. Had they both denied participation in the robbery, their combined sentence would have been four years in prison—two years each. Indeed, cell D offers the lowest combined prison time of any of the outcomes in the payoff matrix. But because the prisoners cannot communicate, each is likely to make a strategic choice that results in a more costly outcome. Of course, the outcome of the game depends on the way the payoff matrix is structured.

FIGURE 11.6 Payoff Matrix for the Prisoners' Dilemma

The four cells represent each of the possible outcomes of the prisoners' game.



Repeated Oligopoly Games

The prisoners' dilemma was played once, by two players. The players were given a payoff matrix; each could make one choice, and the game ended after the first round of choices.

The real world of oligopoly has as many players as there are firms in the industry. They play round after round: a firm raises its price, another firm introduces a new product, the first firm cuts its price, a third firm introduces a new marketing strategy, and so on. An oligopoly game is a bit like a baseball game with an unlimited number of innings—one firm may come out ahead after one round, but another will emerge on top another day. In the computer industry game, the introduction of personal computers changed the rules. IBM, which had won the mainframe game quite handily, struggles to keep up in a world in which rivals continue to slash prices and improve quality.

Oligopoly games may have more than two players, so the games are more complex, but this does not change their basic structure. The fact that the games are repeated introduces new strategic considerations. A player must consider not just the ways in which its choices will affect its rivals now, but how its choices will affect them in the future as well.

We will keep the game simple, however, and consider a duopoly game. The two firms have colluded, either tacitly or overtly, to create a monopoly solution. As long as each player upholds the agreement, the two firms will earn the maximum economic profit possible in the enterprise.

There will, however, be a powerful incentive for each firm to cheat. The monopoly solution may generate the maximum economic profit possible for the two firms combined, but what if one firm captures some of the other firm's profit? Suppose, for example, that two equipment rental firms, Quick Rent and Speedy Rent, operate in a community. Given the economies of scale in the business and the size of the community, it is not likely that another firm will enter. Each firm has about half the market, and they have agreed to charge the prices that would be chosen if the two combined as a single firm. Each earns economic profits of $20,000 per month.

Quick and Speedy could cheat on their arrangement in several ways. One of the firms could slash prices, introduce a new line of rental products, or launch an advertising blitz. This approach would not be likely to increase the total profitability of the two firms, but if one firm could take the other by surprise, it might profit at the expense of its rival, at least for a while.

We will focus on the strategy of cutting prices, which we will call a strategy of cheating on the duopoly agreement. The alternative is not to cheat on the agreement. Cheating increases a firm's profits *if* its rival does not respond. Figure 11.7 shows the payoff matrix facing the two firms at a particular time. As in the prisoners' dilemma matrix, the four cells list the payoffs for the two firms. If neither firm cheats (cell D), profits remain unchanged.

dominant strategy

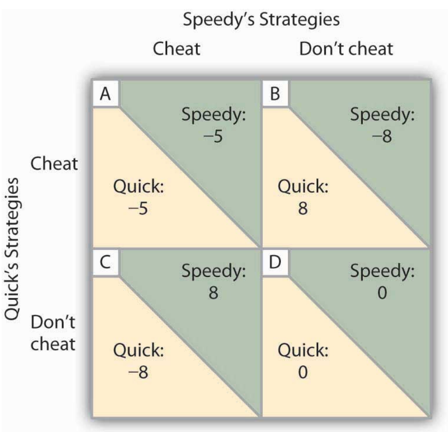
When a player's best strategy is the same regardless of the action of the other player.

dominant strategy equilibrium

A game in which there is a dominant strategy for each player.

FIGURE 11.7 To Cheat or Not to Cheat: Game Theory in Oligopoly

Two rental firms, Quick Rent and Speedy Rent, operate in a duopoly market. They have colluded in the past, achieving a monopoly solution. Cutting prices means cheating on the arrangement; not cheating means maintaining current prices. The payoffs are changes in monthly profits, in thousands of dollars. If neither firm cheats, then neither firm's profits will change. In this game, cheating is a dominant strategy equilibrium.



This game has a dominant strategy equilibrium. Quick's preferred strategy, regardless of what Speedy does, is to cheat. Speedy's best strategy, regardless of what Quick does, is to cheat. The result is that the two firms will select a strategy that lowers their combined profits!

Quick Rent and Speedy Rent face an unpleasant dilemma. They want to maximize profit, yet each is likely to choose a strategy inconsistent with that goal. If they continue the game as it now exists, each will continue to cut prices, eventually driving prices down to the point where price equals average total cost (presumably, the price-cutting will stop there). But that would leave the two firms with zero economic profits.

Both firms have an interest in maintaining the status quo of their collusive agreement. Overt collusion is one device through which the monopoly outcome may be maintained, but that is illegal. One way for the firms to encourage each other not to cheat is to use a tit-for-tat strategy. In a **tit-for-tat strategy** a firm responds to cheating by cheating, and it responds to cooperative behavior by cooperating. As each firm learns that its rival will respond to cheating by cheating, and to cooperation by cooperating, cheating on agreements becomes less and less likely.

Still another way firms may seek to force rivals to behave cooperatively rather than competitively is to use a **trigger strategy**, in which a firm makes clear that it is willing and able to respond to cheating by permanently revoking an agreement. A firm might, for example, make a credible threat to cut prices down to the level of average total cost—and leave them there—in response to any price-cutting by a rival. A trigger strategy is calculated to impose huge costs on any firm that cheats—and on the firm that threatens to invoke the trigger. A firm might threaten to invoke a trigger in hopes that the threat will forestall any cheating by its rivals.

Game theory has proved to be an enormously fruitful approach to the analysis of a wide range of problems. Corporations use it to map out strategies and to anticipate rivals' responses. Governments use it in developing foreign-policy strategies. Military leaders play war games on computers using the basic ideas of game theory. Any situation in which rivals make strategic choices to which competitors will respond can be assessed using game theory analysis.

One rather chilly application of game theory analysis can be found in the period of the Cold War when the United States and the former Soviet Union maintained a nuclear weapons policy that was described by the acronym MAD, which stood for *m*utually *a*ssured *d*estruction. Both countries had enough nuclear weapons to destroy the other several times over, and each threatened to launch sufficient nuclear weapons to destroy the other country if the other country launched a nuclear attack against it or any of its allies. On its face, the MAD doctrine seems, well, mad. It was, after all, a commitment by each nation to respond to any nuclear attack with a counterattack that many scientists expected would end human life on earth. As crazy as it seemed, however, it worked. For 40 years, the two nations did not go to war. While the collapse of the Soviet Union in 1991 ended the need for a MAD doctrine, during the time that the two countries were rivals, MAD was a very effective trigger indeed.

Of course, the ending of the Cold War has not produced the ending of a nuclear threat. Several nations now have nuclear weapons. The threat that Iran will introduce nuclear weapons, given its stated commitment to destroy the state of Israel, suggests that the possibility of nuclear war still haunts the world community.

tit-for-tat strategy

Situation in which a firm responds to cheating by cheating, and responds to cooperative behavior by cooperating.

trigger strategy

Situation in which a firm makes clear that it is willing and able to respond to cheating by permanently revoking an agreement.

KEY TAKEAWAYS

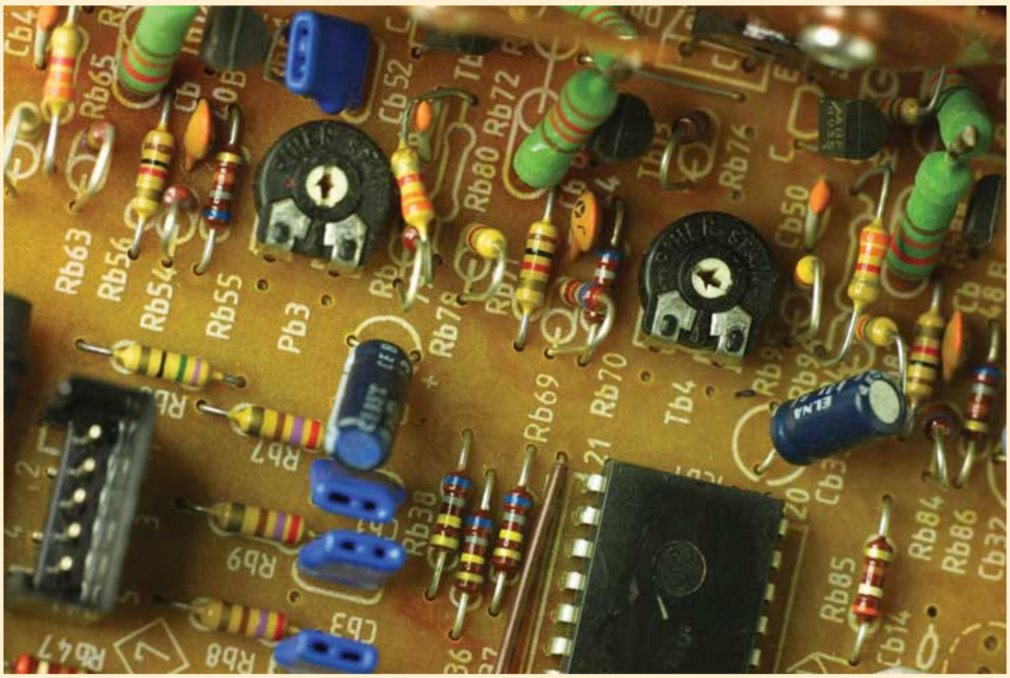
- The key characteristics of oligopoly are a recognition that the actions of one firm will produce a response from rivals and that these responses will affect it. Each firm is uncertain what its rivals' responses might be.
- The degree to which a few firms dominate an industry can be measured using a concentration ratio or a Herfindahl–Hirschman Index.
- One way to avoid the uncertainty firms face in oligopoly is through collusion. Collusion may be overt, as in the case of a cartel, or tacit, as in the case of price leadership.
- Game theory is a tool that can be used to understand strategic choices by firms.
- Firms can use tit-for-tat and trigger strategies to encourage cooperative behavior by rivals.

TRY IT!

Which model of oligopoly would seem to be most appropriate for analyzing firms' behavior in each of the situations given below?

1. When South Airlines lowers its fare between Miami and New York City, North Airlines lowers its fare between the two cities. When South Airlines raises its fare, North Airlines does too.
2. Whenever Bank A raises interest rates on car loans, other banks in the area do too.
3. In 1986, Saudi Arabia intentionally flooded the market with oil in order to punish fellow OPEC members for cheating on their production quotas.
4. In July 1998, Saudi Arabia floated a proposal in which a group of eight or nine major oil-exporting countries (including OPEC members and some nonmembers, such as Mexico) would manage world oil prices by adjusting their production.

Case in Point: Memory Chip Makers Caught in Global Price-Fixing Scheme



© 2010 Jupiterimages Corporation

It may have been the remark by T.L. Chang, vice president of the Taiwan-based memory chip manufacturer Mosel-Vitelic that sparked the investigation by the U.S. Department of Justice Antitrust Division. Mr. Chang was quoted in Taiwan's *Commercial Times* in May 2002 as admitting to price-fixing meetings held in Asia among the major producers of DRAM, or dynamic random access memory. DRAM is the most common semi-conductor main memory format for storage and retrieval of information that is used in personal computers, mobile phones, digital cameras, MP3 music players, and other electronics products. At those meetings, as well as through emails and telephone conferences, the main manufacturers of DRAM decided not only what prices to charge and how much to make available, but also exchanged information on DRAM sales for the purpose of monitoring and enforcing adherence to the agreed prices. The collusion lasted for three years—from 1999 to 2002. In December 2001, DRAM prices were less than $1.00. By May of 2002, price had risen to the $4 to $5 range.

The companies that were directly injured by the higher chip prices included Dell, Compaq, Hewlett-Packard, Apple, IBM, and Gateway. In the end, though, the purchasers of their products paid in the form of higher prices or less memory.

In December 2003, a Micron Technology sales manager pled guilty to obstruction of justice and served six months of home detention. The first chipmaker to plead guilty a year later was Germany-based Infineon Technologies, which was fined $160 million. As of September 2007, five companies, Samsung being the largest, had been charged fines of more than $732 million, and over 3,000 days of jail time had been meted out to eighteen corporate executives.

The sharp reduction in the number of DRAM makers in the late 1990s undoubtedly made it easier to collude. The industry is still quite concentrated with Samsung holding 27.7% of the market and Hynix 21.3%. The price, however, has fallen quite sharply in recent years.

Sources: Department of Justice, "Sixth Samsung Executive Agrees to Plead Guilty to Participating in DRAM Price-Fixing Cartel," Press Release April 19, 2007; Stephen Labaton, "Infineon To Pay a Fine in the Fixing of Chip Prices," The New York Times, September 16, 2004; George Leopold and David Lammers, "DRAMs Under Gun in Antitrust Probe", Electronic Engineering Times, 1124 (June 24, 2002):1, 102; Lee Sun-Young, "Samsung Cements DRAM Leadership," Korea Herald, online, March 31, 2008.

ANSWERS TO TRY IT! PROBLEMS

1. North Airlines seems to be practicing a price strategy known in game theory as tit-for-tat.
2. The banks could be engaged in tacit collusion, with Bank A as the price leader.
3. Saudi Arabia appears to have used a trigger strategy, another aspect of game theory. In general, of course, participants hope they will never have to "pull" the trigger, because doing so harms all participants. After years of cheating by other OPEC members, Saudi Arabia did undertake a policy that hurt all members of OPEC, including itself; OPEC has never since regained the prominent role it played in oil markets.
4. Saudi Arabia seems to be trying to create another oil cartel, a form of overt collusion.

3. EXTENSIONS OF IMPERFECT COMPETITION: ADVERTISING AND PRICE DISCRIMINATION

LEARNING OBJECTIVES

1. **Discuss the possible effects of advertising on competition, price, and output.**
2. **Define price discrimination, list the conditions that make it possible, and explain the relationship between the price charged and price elasticity of demand.**

The models of monopoly and of imperfectly competitive markets allow us to explain two commonly observed features of many markets: advertising and price discrimination. Firms in markets that are not perfectly competitive try to influence the positions of the demand curves they face, and hence profits, through advertising. Profits may also be enhanced by charging different customers different prices. In this section we will discuss these aspects of the behavior of firms in markets that are not perfectly competitive.

3.1 Advertising

Firms in monopoly, monopolistic competition, and oligopoly use advertising when they expect it to increase their profits. We see the results of these expenditures in a daily barrage of advertising on television, radio, newspapers, magazines, billboards, passing buses, park benches, the mail, home telephones, and the ubiquitous pop-up advertisements on our computers—in virtually every medium imaginable. Is all this advertising good for the economy?

We have already seen that a perfectly competitive economy with fully defined and easily transferable property rights will achieve an efficient allocation of resources. There is no role for advertising in such an economy, because everyone knows that firms in each industry produce identical products. Furthermore, buyers already have complete information about the alternatives available to them in the market.

But perfect competition contrasts sharply with imperfect competition. Imperfect competition can lead to a price greater than marginal cost and thus generate an inefficient allocation of resources. Firms in an imperfectly competitive market may advertise heavily. Does advertising cause inefficiency, or is it part of the solution? Does advertising insulate imperfectly competitive firms from competition and allow them to raise their prices even higher, or does it encourage greater competition and push prices down?

There are two ways in which advertising could lead to higher prices for consumers. First, the advertising itself is costly; in 2007, firms in the United States spent about $149 billion on advertising. By pushing up production costs, advertising may push up prices. If the advertising serves no socially useful purpose, these costs represent a waste of resources in the economy. Second, firms may be able to use advertising to manipulate demand and create barriers to entry. If a few firms in a particular market have developed intense brand loyalty, it may be difficult for new firms to enter—the advertising creates a kind of barrier to entry. By maintaining barriers to entry, firms may be able to sustain high prices.

But advertising has its defenders. They argue that advertising provides consumers with useful information and encourages price competition. Without advertising, these defenders argue, it would be impossible for new firms to enter an industry. Advertising, they say, promotes competition, lowers prices, and encourages a greater range of choice for consumers.

Advertising, like all other economic phenomena, has benefits as well as costs. To assess those benefits and costs, let us examine the impact of advertising on the economy.

Advertising and Information

Advertising does inform us about products and their prices. Even critics of advertising generally agree that when advertising advises consumers about the availability of new products, or when it provides price information, it serves a useful function. But much of the information provided by advertising appears to be of limited value. Hearing that "Pepsi is the right one, baby" or "Tide gets your clothes whiter than white" may not be among the most edifying lessons consumers could learn.

Some economists argue, however, that even advertising that seems to tell us nothing may provide useful information. They note that a consumer is unlikely to make a repeat purchase of a product that turns out to be a dud. Advertising an inferior product is likely to have little payoff; people who do try it are not likely to try it again. It is not likely a firm could profit by going to great expense to launch a product that produced only unhappy consumers. Thus, if a product is heavily advertised, its producer is likely to be confident that many consumers will be satisfied with it and make repeat purchases. If this is the case, then the fact that the product is advertised, regardless of the content of that advertising, signals consumers that at least its producer is confident that the product will satisfy them.

Advertising and Competition

If advertising creates consumer loyalty to a particular brand, then that loyalty may serve as a barrier to entry to other firms. Some brands of household products, such as laundry detergents, are so well established they may make it difficult for other firms to enter the market.

In general, there is a positive relationship between the degree of concentration of market power and the fraction of total costs devoted to advertising. This relationship, critics argue, is a causal one; the high expenditures on advertising are the cause of the concentration. To the extent that advertising increases industry concentration, it is likely to result in higher prices to consumers and lower levels of output. The higher prices associated with advertising are not simply the result of passing on the cost of the advertising itself to consumers; higher prices also derive from the monopoly power the advertising creates.

But advertising may encourage competition as well. By providing information to consumers about prices, for example, it may encourage price competition. Suppose a firm in a world of no advertising wants to increase its sales. One way to do that is to lower price. But without advertising, it is extremely difficult to inform potential customers of this new policy. The likely result is that there would be little response, and the price experiment would probably fail. Price competition would thus be discouraged in a world without advertising.

Empirical studies of markets in which advertising is not allowed have confirmed that advertising encourages price competition. One of the most famous studies of the effects of advertising looked at pricing for prescription eyeglasses. In the early 1970s, about half the states in the United States banned advertising by firms making prescription eyeglasses; the other half allowed it. A comparison of prices in the two groups of states by economist Lee Benham showed that the cost of prescription eyeglasses was far lower in states that allowed advertising than in states that banned it.[4] Mr. Benham's research proved quite influential—virtually all states have since revoked their bans on such advertising. Similarly, a study of the cigarette industry revealed that before the 1970 ban on radio and television advertising market shares of the leading cigarette manufacturers had been declining, while after the ban market shares and profit margins increased.[5]

Advertising may also allow more entry by new firms. When Kia, a South Korean automobile manufacturer, entered the U.S. low-cost compact car market in 1994, it flooded the airwaves with advertising. Suppose such advertising had not been possible. Could Kia have entered the market in the United States? It seems highly unlikely that any new product could be launched without advertising. The absence of advertising would thus be a barrier to entry that would increase the degree of monopoly power in the economy. A greater degree of monopoly power would, over time, translate into higher prices and reduced output.

Advertising is thus a two-edged sword. On the one hand, the existence of established and heavily advertised rivals may make it difficult for a new firm to enter a market. On the other hand, entry into most industries would be virtually impossible without advertising.

Economists do not agree on whether advertising helps or hurts competition in particular markets, but one general observation can safely be made—a world with advertising is more competitive than a world without advertising would be. The important policy question is more limited—and more difficult to answer: Would a world with *less* advertising be more competitive than a world with more?

3.2 Price Discrimination

Throughout the text up to this point, we have assumed that firms sold all units of output at the same price. In some cases, however, firms can charge different prices to different consumers. If such an opportunity exists, the firm can increase profits further.

price discrimination

Situation in which a firm charges different prices for the same good or service to different consumers, even though there is no difference in the cost to the firm of supplying these consumers.

When a firm charges different prices for the same good or service to different consumers, even though there is no difference in the cost to the firm of supplying these consumers, the firm is engaging in **price discrimination**. Except for a few situations of price discrimination that have been declared illegal, such as manufacturers selling their goods to distributors at different prices when there are no differences in cost, price discrimination is generally legal.

The potential for price discrimination exists in all market structures except perfect competition. As long as a firm faces a downward-sloping demand curve and thus has some degree of monopoly power, it may be able to engage in price discrimination. But monopoly power alone is not enough to allow a firm to price discriminate. Monopoly power is one of three conditions that must be met:

1. **A Price-Setting Firm** The firm must have some degree of monopoly power—it must be a price setter. A price-taking firm can only take the market price as given—it is not in a position to make price choices of any kind. Thus, firms in perfectly competitive markets will not engage in price discrimination. Firms in monopoly, monopolistically competitive, or oligopolistic markets may engage in price discrimination.
2. **Distinguishable Customers** The market must be capable of being fairly easily segmented—separated so that customers with different elasticities of demand can be identified and treated differently.
3. **Prevention of Resale** The various market segments must be isolated in some way from one another to prevent customers who are offered a lower price from selling to customers who are charged a higher price. If consumers can easily resell a product, then discrimination is unlikely to be successful. Resale may be particularly difficult for certain services, such as dental checkups.

Examples of price discrimination abound. Senior citizens and students are often offered discount fares on city buses. Children receive discount prices for movie theater tickets and entrance fees at zoos and theme parks. Faculty and staff at colleges and universities might receive discounts at the campus bookstore. Airlines give discount prices to customers who are willing to stay over a Saturday night. Physicians might charge wealthy patients more than poor ones. People who save coupons are able to get discounts on many items. In all these cases a firm charges different prices to different customers for what is essentially the same product.

Not every instance of firms charging different prices to different customers constitutes price discrimination. Differences in prices may reflect different costs associated with providing the product. One buyer might require special billing practices, another might require delivery on a particular day of the week, and yet another might require special packaging. Price differentials based on differences in production costs are not examples of price discrimination.

Why would a firm charge different prices to different consumers? The answer can be found in the marginal decision rule and in the relationship between marginal revenue and elasticity.

Suppose an airline has found that its long-run profit-maximizing solution for a round-trip flight between Minneapolis and Cleveland, when it charges the same price to all passengers, is to carry 300 passengers at $200 per ticket. The airline has a degree of monopoly power, so it faces a downward-sloping demand curve. The airline has noticed that there are essentially two groups of customers on each flight: people who are traveling for business reasons and people who are traveling for personal reasons (visiting family or friends or taking a vacation). We will call this latter group "tourists." Of the 300 passengers, 200 are business travelers and 100 are tourists. The airline's revenue from business travelers is therefore currently $40,000 ($200 times 200 business travelers) and from tourists is currently $20,000 ($200 times 100 tourists).

It seems likely that the price elasticities of demand of these two groups for a particular flight will differ. Tourists may have a wide range of substitutes: They could take their trips at a different time, they could vacation in a different area, or they could easily choose not to go at all. Business travelers, however, might be attending meetings or conferences at a particular time and in a particular city. They have options, of course, but the range of options is likely to be more limited than the range of options facing tourists. Given all this, tourists are likely to have relatively more price elastic demand than business travelers for a particular flight.

The difference in price elasticities suggests the airline could increase its profit by adjusting its pricing. To simplify, suppose that at a price of about $200 per ticket, demand by tourists is relatively price elastic and by business travelers is relatively less price elastic. It is plausible that the marginal cost of additional passengers is likely to be quite low, since the number of crewmembers will not vary and no food is served on short flights. Thus, if the airline can increase its revenue, its profits will increase. Suppose the airline lowers the price for tourists to $190. Suppose that the lower price encourages 10 more tourists to take the flight. Of course, the airline cannot charge different prices to different tourists; rather it charges $190 to all, now 110, tourists. Still, the airline's revenue from tourist passengers increases from $20,000 to $20,900 ($190 times 110 tourists). Suppose it charges $250 to its business travelers. As a result, only 195 business travelers take the flight. The airline's revenue from business travelers still rises from $40,000 to $48,750 ($250 times 195 business travelers). The airline will continue to change the mix of passengers, and increase the number of passengers, so long as doing so increases its profit. Because tourist demand is relatively price elastic, relatively small reductions in price will attract relatively large numbers of additional tourists. Because business demand is relatively less elastic, relatively large increases in price will discourage relatively small numbers of business travelers from making the trip. The airline will continue to reduce its price to tourists and raise its price to business travelers as long as it gains profit from doing so.

Of course, the airline can impose a discriminatory fare structure only if it can distinguish tourists from business travelers. Airlines typically do this by looking at the travel plans of their customers. Trips that involve a stay over a weekend, for example, are more likely to be tourist related, whereas trips that begin and end during the workweek are likely to be business trips. Thus, airlines charge much lower fares for trips that extend through a weekend than for trips that begin and end on weekdays.

In general, price-discrimination strategies are based on differences in price elasticity of demand among groups of customers and the differences in marginal revenue that result. A firm will seek a price structure that offers customers with more elastic demand a lower price and offers customers with relatively less elastic demand a higher price.

It is always in the interest of a firm to discriminate. Yet most of the goods and services that we buy are not offered on a discriminatory basis. A grocery store does not charge a higher price for vegetables to vegetarians, whose demand is likely to be less elastic than that of its omnivorous customers. An audio store does not charge a different price for Pearl Jam's compact disks to collectors seeking a complete collection than it charges to casual fans who could easily substitute a disk from another performer. In these cases, firms lack a mechanism for knowing the different demands of their customers and for preventing resale.

KEY TAKEAWAYS

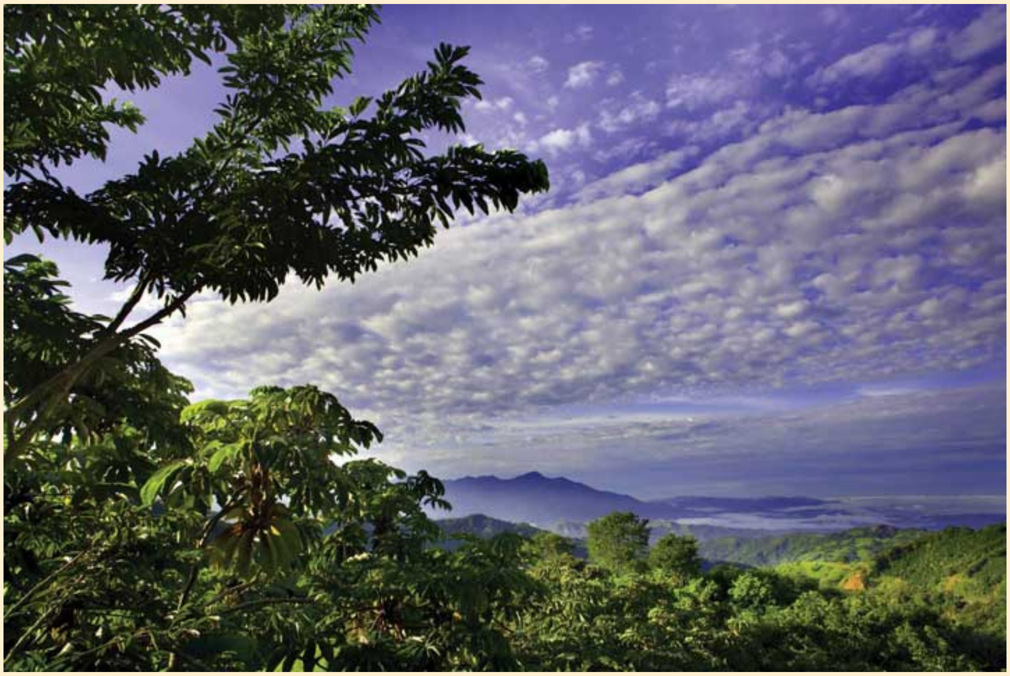
- If advertising reduces competition, it tends to raise prices and reduce quantities produced. If it enhances competition, it tends to lower prices and increase quantities produced.
- In order to engage in price discrimination, a firm must be a price setter, must be able to identify consumers whose elasticities differ, and must be able to prevent resale of the good or service among consumers.
- The price-discriminating firm will adjust its prices so that customers with more elastic demand pay lower prices than customers with less elastic demand.

TRY IT!

Explain why price discrimination is often found in each of the following settings. Does it make sense in terms of price elasticity of demand?

1. Senior citizen discounts for travel
2. Food sold cheaper if the customer has a coupon for the item
3. College scholarships to students with the best academic records or to students with special athletic, musical, or other skills

Case in Point: Pricing Costa Rica's National Parks



© 2010 Jupiterimages Corporation

Costa Rica boasts some of the most beautiful national parks in the world. An analysis by Francisco Alpizar, an economist with Gothenburg University in Sweden and CATIE, a tropical research institute in Costa Rica, suggests that Costa Rica should increase the degree to which it engages in price discrimination in pricing its national parks.

The country has experimented with a wide range of prices for its national parks, with the price varying between $.80 and $15 for a daily visit. With data on the resultant number of visitors at each price, Professor Alpizar was able to estimate the demand curve, compute the price elasticity of demand, and develop a recommendation for pricing the country's national parks.

Presumably, foreign visitors have a relatively less elastic demand for visiting the parks than do local citizens. Local citizens have better knowledge of substitutes for the parks—namely other areas in Costa Rica. And, of course, once foreign travelers are in the country, they have already committed the expense of getting there, and are less likely to be willing to pass up a visit to national parks based on pricing considerations.

Costa Rica already discriminates to a large degree. Foreigners are charged $7 per day to visit the parks; locals are charged $2. Professor Alpizar proposes increasing the degree of discrimination.

He estimates that the price elasticity of foreign demand for visits to Costa Rica's national parks is −0.68. That, of course, suggests inelastic demand. Costa Rica could increase its revenue from foreign visitors by increasing the fee. Professor Alpizar proposes increasing the fee for foreigners to $10. He proposes that the price charged to Costa Ricans remain at $2—a price that he calculates equals the marginal cost of an additional visit.

Professor Alpizar calculates a fee of $10 per visit by a foreigner would more than pay the country's fixed cost of maintaining its extensive park system, which utilizes 24% of the country's land. The higher price would thus allow the government to meet the major costs of operating the national parks. Charging a $2 fee to locals would satisfy the efficiency requirement that price equal marginal cost for local visitors; the $10 fee to foreigners would permit the country to exploit its monopoly power in permitting people to visit the parks. The Costa Rican government has asked Professor Alpizar to design three pilot projects aimed at incorporating his proposal to raise park fees to foreign visitors.

Source: Francisco Alpizar, "The Pricing of Protected Areas in Nature-Based Tourism: A Local Prospective," Ecological Economics, 56(2) (February 2006): 294–307 and personal correspondence with Professor Alpizar.

ANSWERS TO TRY IT! PROBLEMS

1. Senior citizens are (usually!) easy to identify, and for travel, preventing resale is usually quite easy as well. For example, a picture ID is required to board an airplane. Airlines might be expected to oppose implementing the rule since it is costly for them. The fact that they support the rule can be explained by how it aids them in practicing price discrimination, by preventing the resale of discount tickets, which now can easily be matched to the purchasing customers. The demand for air travel by senior citizens is likely to be more elastic than it is for other passengers, especially business travelers, since the purpose of their travel is largely discretionary (often touristic in nature) and since their time is likely to be less costly, making them more willing to seek out information on travel alternatives than the rest of the population.
2. Since the customer must present the coupon at the point of sale, identification is easy. Willingness to search for and cut out coupons suggests a high degree of price consciousness and thus a greater price elasticity of demand.
3. Such students are likely to have more choices of where to attend college. As we learned in an earlier chapter on elasticity, demand is likely to be more elastic when substitutes are available for it. Enrollment procedures make identification and prevention of resale very easy.

4. REVIEW AND PRACTICE

Summary

This chapter examined the world of imperfect competition that exists between the idealized extremes of perfect competition and monopoly. Imperfectly competitive markets exist whenever there is more than one seller in a market and at least one seller has some degree of control over price.

We discussed two general types of imperfectly competitive markets: monopolistic competition and oligopoly. Monopolistic competition is characterized by many firms producing similar but differentiated goods and services in a market with easy entry and exit. Oligopoly is characterized by relatively few firms producing either standardized or differentiated products. There may be substantial barriers to entry and exit.

In the short run, a monopolistically competitive firm's pricing and output decisions are the same as those of a monopoly. In the long run, economic profits will be whittled away by the entry of new firms and new products that increase the number of close substitutes. An industry dominated by a few firms is an oligopoly. Each oligopolist is aware of its interdependence with other firms in the industry and is constantly aware of the behavior of its rivals. Oligopolists engage in strategic decision making in order to determine their best output and pricing strategies as well as the best forms of nonprice competition.

Advertising in imperfectly competitive markets can increase the degree of competitiveness by encouraging price competition and promoting entry. It can also decrease competition by establishing brand loyalty and thus creating barriers to entry.

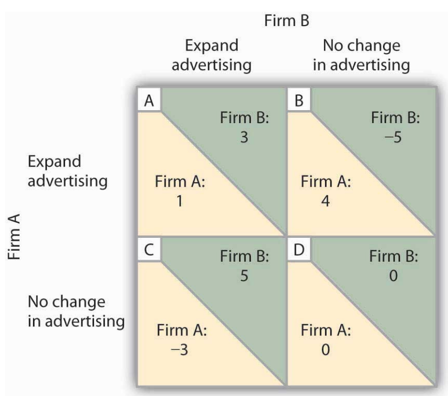
Where conditions permit, a firm can increase its profits by price discrimination, charging different prices to customers with different elasticities of demand. To practice price discrimination, a price-setting firm must be able to segment customers that have different elasticities of demand and must be able to prevent resale among its customers.

CONCEPT PROBLEMS

1. What are the major distinctions between a monopolistically competitive industry and an oligopolistic industry?
2. What is the difference between a price taker and a price setter? Which do you think a firm would prefer to be? Why?
3. In the model of monopolistic competition, we say that there is product differentiation. What does this mean, and how does it differ from the assumption of homogeneous goods in perfect competition?
4. In the following list of goods and services, determine whether the item is produced under conditions of monopolistic competition or of oligopoly.
    a. soft drinks
    b. exercise drinks
    c. office supply stores
    d. massage therapists
    e. accountants
    f. colleges and universities
    g. astrologists
5. Suppose a city experiences substantial population growth. What is likely to happen to profits in the short run and in the long run in the market for haircuts, a monopolistically competitive market?
6. Some professors grade students on the basis of an absolute percentage of the highest score earned on each test given during the semester. All students who get within a certain percentage of the highest score earned get an A. Why do these professors not worry that the students will get together and collude in such a way as to keep the high score in the class equal to a very low total?
7. Your parents probably told you to avoid tit-for-tat behavior. Why does it make sense for firms to do it?
8. What model of oligopoly behavior were the DRAM producers discussed in the Case in Point following? How might the DRAM producers have achieved their goal and still stayed within the law?
9. Explain why a price increase for foreigners would increase Costa Rica's total revenue and profits from operating its national park system.
10. Restaurants typically charge much higher prices for dinner than for lunch, despite the fact that the cost of serving these meals is about the same. Why do you think this is the case? (*Hint:* Think about the primary consumers of these meals and their respective elasticities.)
11. What effect do you think advertising to discourage cigarette smoking will have on teens? On adults? What changes might occur in the cigarette market as a result?
12. Many manufacturers of clothing and other consumer goods open stores in outlet malls where they charge much lower prices than they charge in their own stores located within cities. Outlet malls are typically located a considerable distance from major metropolitan areas, and stores in them typically charge much lower prices than do stores located within cities. Given that both sets of stores are often owned by the same firm, explain this price discrimination based on likely differences in the price elasticity of demand between consumers in the two types of stores.
13. Suppose a particular state were to ban the advertising of prices charged by firms that provide laser eye surgery. What effect do you think that would have on the price of this service?
14. The Case in Point on microbreweries noted that a large number of such breweries open every year. Yet, the model of monopolistic competition predicts that the long run equilibrium solution in such markets is one of zero economic profits. Why do firms enter such industries?
15. Many lawyers advertise their services. Do you think this raises or lowers the price of legal services? Explain your answer carefully.

NUMERICAL PROBLEMS

1. Suppose the monopolistically competitive barber shop industry in a community is in long-run equilibrium, and that the typical price is $20 per haircut. Moreover, the population is rising.
    a. Illustrate the short-run effects of a change on the price and output of a typical firm in the market.
    b. Show what happens in the long run. Will the final price be higher than $20? Equal $20? Be less than $20? Assume that nothing happens to the cost of producing haircuts.
    c. Suppose that, initially, the price of a typical children's haircut is $10. Do you think this represents price discrimination? Why or why not?
2. Consider the same industry as in Problem 1. Suppose the market is in long-run equilibrium and that an annual license fee is imposed on barber shops.
    a. Illustrate the short-run effects of the change on the price and output of haircuts for a typical firm in the community.
    b. Now show what happens to price and output for a typical firm in the long run.
    c. Who pays the fee in the long run? How does this compare to the conclusions of the model of perfect competition?
3. Industry A consists of four firms, each of which has an equal share of the market.
    a. Compute the Herfindahl-Hirschman index for the industry.
    b. Industry B consists of 10 firms, each of which has an equal share of the market. Compare the Herfindahl–Hirschman Indexes for the two industries.
    c. Now suppose that there are 100 firms in the industry, each with equal shares. What is the Herfindahl-Hirschman index for this industry?
    d. State the general relationship between the competitiveness of an industry and its Herfindahl-Hirschman index.
4. Given the payoff matrix (shown below) for a duopoly, consisting of Firm A and Firm B, in which each firm is considering an expanded advertising campaign, answer the following questions (all figures in the payoff matrix give changes in annual profits in millions of dollars):
    a. Does Firm A have a dominant strategy?
    b. Does Firm B have a dominant strategy?
    c. Is there a dominant strategy equilibrium? Explain.



5. Suppose that two industries each have a four-firm concentration ratio of 75%.
    a. Explain what this means.
    b. Suppose that the HHI of the first industry is 425, and that the HHI of the second is 260. Which would you say is the more competitive? Why?
6. Suppose that a typical firm in a monopolistically competitive industry faces a demand curve given by:

$$q = 60 - (1/2)p, \text{ where } q \text{ is quantity sold per week.}$$

The firm's marginal cost curve is given by: $MC = 60$.

    a. How much will the firm produce in the short run?
    b. What price will it charge?
    c. Draw the firm's demand, marginal revenue, and marginal cost curves. Does this solution represent a long-run equilibrium? Why or why not?

ENDNOTES

1. Victoria Murphy Barrett, "Reading Your Mind," *Forbes Online*, October 29, 2007, 180 Issue 9, p. 50.
2. Mylene Mangalindan and Robert A. Guth, "eBay Talks to Microsoft, Yahoo About a Common Foe: Google," *The Wall Street Journal Online*, April 21, 2006, p. A1.
3. One legal cartel is the NCAA, which many economists regard as a successful device through which member firms (colleges and universities) collude on a wide range of rules through which they produce sports.
4. Lee Benham, "The Effect of Advertising on the Price of Eyeglasses," *Journal of Law and Economics* 15(2) (1972): 337–352.
5. Woodrow Eckard, "Competition and the Cigarette TV Advertising Ban," *Economic Inquiry* 29(1) (January 1991), 119–133.

CHAPTER 15

Public Finance and Public Choice

START UP: WHERE YOUR TAX DOLLARS GO

You pay sales taxes on most of the goods you purchase. If you smoke or drink or drive a car, you pay taxes on cigarettes, alcohol, and gasoline. If you work, you may pay income and payroll taxes.

What does the government do with the taxes it collects? If you go to a public school, you are a consumer of public sector services. You also consume the services of the public sector when you drive on a public street or go to a public park. You consume public sector services since you are protected by law enforcement agencies and by the armed forces. And the production of everything else you consume is affected by regulations imposed by local, state, or federal agencies.

The public sector is a crucially important segment of the economy, due in part to its size. The nearly 90,000 government jurisdictions in the United States, from local fire protection districts to the federal government, either produce or purchase nearly one-fifth of all domestic goods and services. The U.S. government is the largest single purchaser of goods and services in the world.

This chapter examines the role of government in a market economy and the ways in which the taxes that support government affect economic behavior. The study of government expenditure and tax policy and of their impact on the economy is called **public finance**.

public finance

The study of government expenditure and tax policy and of their impact on the economy.

We will also explore the economics of public sector choices. Economists put the notions of self-interest and the marginal decision rule to work in the analysis of choices made by people in the public sector—voters, government employees, interest groups, and politicians.

1. THE ROLE OF GOVERNMENT IN A MARKET ECONOMY

LEARNING OBJECTIVES

1. **Discuss and illustrate government responses to the market failures of public goods, external costs and benefits, and imperfect competition and how these responses have the potential to reduce deadweight loss.**
2. **Define merit and demerit goods and explain why government may intervene to affect the quantities consumed.**
3. **Discuss ways in which governments redistribute income.**

What do we want from our government? One answer is that we want a great deal more than we did several decades ago. The role of government has expanded dramatically in the last 75+ years. In 1929 (the year the Commerce Department began keeping annual data on macroeconomic performance in the United States), government expenditures at all levels (state, local, and federal) were less than 10% of the nation's total output, which is called gross domestic product (GDP). In the current century, that share has more than tripled. Total government spending per capita, adjusted for inflation, has increased more than six fold since 1929.

government expenditures

All spending by government agencies.

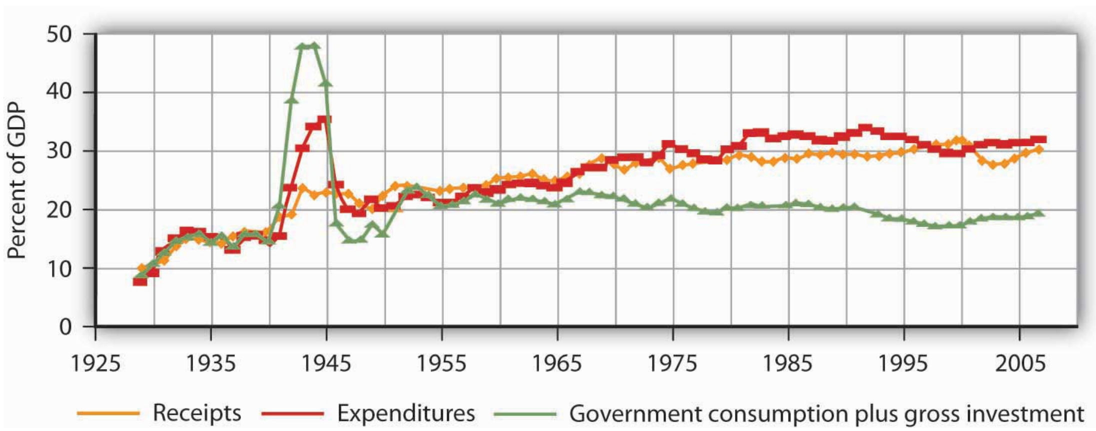
government revenues

All funds received by government agencies.

Figure 15.1 shows total government expenditures and revenues as a percentage of GDP from 1929 to 2007. All levels of government are included. **Government expenditures** include all spending by government agencies. **Government revenues** include all funds received by government agencies. The primary component of government revenues is taxes; revenue also includes miscellaneous receipts from fees, fines, and other sources. We will look at types of government revenues and expenditures later in this chapter.

FIGURE 15.1 Government Expenditures and Revenues as a Percentage of GDP

Government expenditures and revenues have risen dramatically as a percentage of GDP, the most widely used measure of economic activity.



Source: U.S. Department of Commerce, Bureau of Economic Analysis, NIPA Tables 1.15 and 3.1.

government purchases

Goods or services purchased by a government agency.

transfer payments

Payments made by government agencies to individuals in the form of grants rather than in return for labor or other services.

Figure 15.1 also shows government purchases as a percentage of GDP. **Government purchases** happen when a government agency purchases or produces a good or a service. We measure government purchases to suggest the opportunity cost of government. Whether a government agency purchases a good or service or produces it, factors of production are being used for public sector, rather than private sector, activities. A city police department's purchase of new cars is an example of a government purchase. Spending for public education is another example.

Government expenditures and purchases are not equal because much government spending is not for the purchase of goods and services. The primary source of the gap is **transfer payments**, payments made by government agencies to individuals in the form of grants rather than in return for labor or other services. Transfer payments represent government expenditures but not government purchases. Governments engage in transfer payments in order to redistribute income from one group to another. The various welfare programs for low-income people are examples of transfer payments. Social Security is the largest transfer payment program in the United States. This program transfers income from people who are working (by taxing their pay) to people who have retired. Interest payments on government debt, which are also a form of expenditure, are another example of an expenditure that is not counted as a government purchase.

Several points about Figure 15.1 bear special attention. Note first the path of government purchases. Government purchases relative to GDP rose dramatically during World War II, then dropped back to about their prewar level almost immediately afterward. Government purchases rose again, though less sharply, during the Korean War. This time, however, they did not drop back very far after the war. It was during this period that military spending rose to meet the challenge posed by the former Soviet Union and other communist states—the "Cold War." Government purchases have ranged between 15 and 20% of GDP ever since. The Vietnam War, the Persian Gulf War, and the wars in Afghanistan and Iraq did not have the impact on purchases that characterized World War II or even the Korean War. A second development, the widening gap between expenditures and purchases, has occurred since the 1960s. This reflects the growth of federal transfer programs, principally Social Security, programs to help people pay for health-care costs, and aid to low-income people. We will discuss these programs later in this chapter.

Finally, note the relationship between expenditures and receipts. When a government's revenues equal its expenditures for a particular period, it has a **balanced budget**. A **budget surplus** occurs if a government's revenues exceed its expenditures, while a **budget deficit** exists if government expenditures exceed revenues.

balanced budget

Situation that occurs when a government's revenues equal its expenditures for a particular period.

budget surplus

Situation that occurs when a government's revenues exceed its expenditures.

budget deficit

Situation that occurs when government expenditures exceed revenues.

Prior to 1980, revenues roughly matched expenditures for the public sector as a whole, except during World War II. But expenditures remained consistently higher than revenues between 1980 and 1996. The federal government generated very large deficits during this period, deficits that exceeded surpluses that typically occur at the state and local levels of government. The largest increases in spending came from Social Security and increased health-care spending at the federal level. Efforts by the federal government to reduce and ultimately eliminate its deficit, together with surpluses among state and local governments, put the combined budget for the public sector in surplus beginning in 1997. As of 1999, the Congressional Budget Office was predicting that increased federal revenues produced by a growing economy would continue to produce budget surpluses well into the twenty-first century.

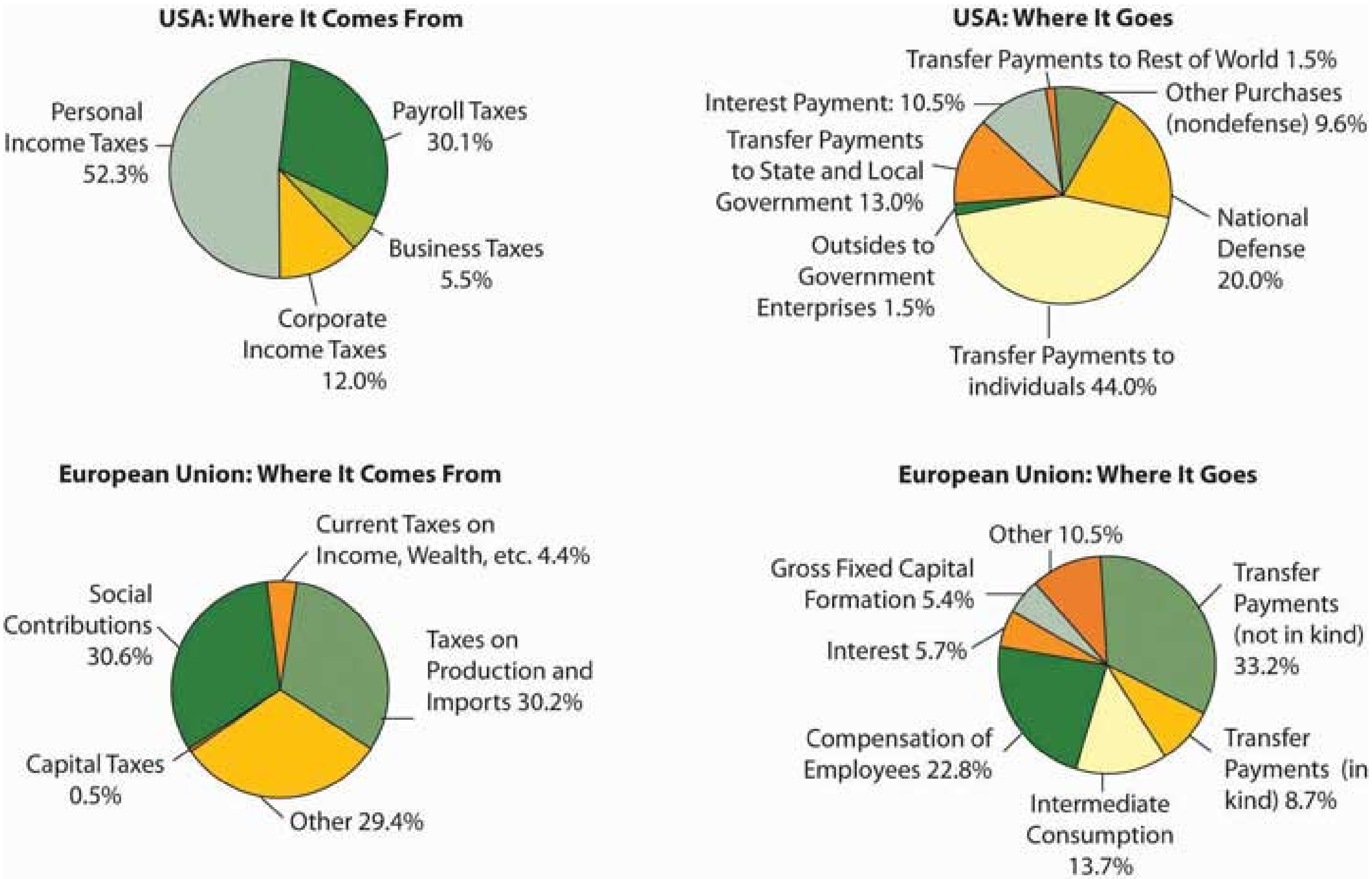
That rather rosy forecast was set aside after September 11, 2001. Terrorist attacks on the United States and later on several other countries led to sharp and sustained increases in federal spending for wars in Afghanistan and Iraq, as well as expenditures for Homeland Security. The administration of George W. Bush proposed, and Congress approved, a tax cut. The combination of increased spending on the abovementioned items and others, as well as tax cuts, produced substantial deficits.

The evidence presented in Figure 15.1 does not fully capture the rise in demand for public sector services. In addition to governments that spend more, people in the United States have clearly chosen governments that do more. The scope of regulatory activity conducted by governments at all levels, for example, has risen sharply in the last several decades. Regulations designed to prevent discrimination, to protect consumers, and to protect the environment are all part of the response to a rising demand for public services, as are federal programs in health care and education.

Figure 15.2 summarizes the main revenue sources and types of expenditures for the U.S. federal government and for the European Union. In the United States, most revenues came from personal income taxes and from payroll taxes. Most expenditures were for transfer payments to individuals. Federal purchases were primarily for national defense; the "other purchases" category includes things such as spending for transportation projects and for the space program. Interest payments on the national debt and grants by the federal government to state and local governments were the other major expenditures. The situation in the European Union differs primarily by the fact that a greater share of revenue comes from taxes on production and imports and substantially less is spent on defense.

FIGURE 15.2 Government Revenue Sources and Expenditures: 2007

The four panels show the sources of government revenues and the shares of expenditures on various activities for all levels of government in the United States and the European Union in 2007.



To understand the role of government, it will be useful to distinguish four broad types of government involvement in the economy. First, the government attempts to respond to market failures to allocate resources efficiently. In a particular market, efficiency means that the quantity produced is determined by the intersection of a demand curve that reflects all the benefits of consuming a particular good or service and a supply curve that reflects the opportunity costs of producing it. Second, government agencies act to encourage or discourage the consumption of certain goods and services. The prohibition of drugs such as heroin and cocaine is an example of government seeking to discourage consumption of these drugs. Third, the government redistributes income through programs such as welfare and Social Security. Fourth, the government can use its spending and tax policies to influence the level of economic activity and the price level.

We will examine the first three of these aspects of government involvement in the economy in this chapter. The fourth, efforts to influence the level of economic activity and the price level, fall within the province of macroeconomics.

1.1 Responding to Market Failure

In an earlier chapter on markets and efficiency, we learned that a market maximizes net benefit by achieving a level of output at which marginal benefit equals marginal cost. That is the efficient solution. In most cases, we expect that markets will come close to achieving this result—that is the important lesson of Adam Smith's idea of the market as an invisible hand, guiding the economy's scarce factors of production to their best uses. That is not always the case, however.

We have studied several situations in which markets are unlikely to achieve efficient solutions. In an earlier chapter, we saw that private markets are likely to produce less than the efficient quantities of public goods such as national defense. They may produce too much of goods that generate external costs and too little of goods that generate external benefits. In cases of imperfect competition, we have seen that the market's output of goods and services is likely to fall short of the efficient level. In all these cases, it is possible that government intervention will move production levels closer to their efficient quantities. In the next three sections, we shall review how a government could improve efficiency in the cases of public goods, external costs and benefits, and imperfect competition.

Public Goods

A public good is a good or service for which exclusion is prohibitively costly and for which the marginal cost of adding another consumer is zero. National defense, law enforcement, and generally available knowledge are examples of public goods.

The difficulty posed by a public good is that, once it is produced, it is freely available to everyone. No consumer can be excluded from consumption of the good on grounds that he or she has not paid for it. Consequently, each consumer has an incentive to be a free rider in consuming the good, and the firms providing a public good do not get a signal from consumers that reflects their benefit of consuming the good.

Certainly we can expect some benefits of a public good to be revealed in the market. If the government did not provide national defense, for example, we would expect some defense to be produced, and some people would contribute to its production. But because free-riding behavior will be common, the market's production of public goods will fall short of the efficient level.

The theory of public goods is an important argument for government involvement in the economy. Government agencies may either produce public goods themselves, as do local police departments, or pay private firms to produce them, as is the case with many government-sponsored research efforts. An important debate in the provision of public education revolves around the question of whether education should be produced by the government, as is the case with traditional public schools, or purchased by the government, as is done in charter schools.

External Costs and Benefits

External costs are imposed when an action by one person or firm harms another, outside of any market exchange. The **social cost** of producing a good or service equals the private cost plus the external cost of producing it. In the case of external costs, private costs are less than social costs.

Similarly, external benefits are created when an action by one person or firm benefits another, outside of any market exchange. The **social benefit** of an activity equals the private benefit revealed in the market plus external benefits. When an activity creates external benefits, its social benefit will be greater than its private benefit.

The lack of a market transaction means that the person or firm responsible for the external cost or benefit does not face the full cost or benefit of the choice involved. We expect markets to produce more than the efficient quantity of goods or services that generate external costs and less than the efficient quantity of goods or services that generate external benefits.

Consider the case of firms that produce memory chips for computers. The production of these chips generates water pollution. The cost of this pollution is an external cost; the firms that generate it do not face it. These firms thus face some, but not all, of the costs of their production choices. We can expect the market price of chips to be lower, and the quantity produced greater, than the efficient level.

Inoculations against infectious diseases create external benefits. A person getting a flu shot, for example, receives private benefits; he or she is less likely to get the flu. But there will be external benefits as well: Other people will also be less likely to get the flu because the person getting the shot is less likely to have the flu. Because this latter benefit is external, the social benefit of flu shots exceeds the private benefit, and the market is likely to produce less than the efficient quantity of flu shots. Public, private, and charter schools often require such inoculations in an effort to get around the problem of external benefits.

social cost

The private cost of producing a good or service plus the external cost of producing it.

social benefit

The private benefit of a good revealed in the market plus external benefits.

Imperfect Competition

In a perfectly competitive market, price equals marginal cost. If competition is imperfect, however, individual firms face downward-sloping demand curves and will charge prices greater than marginal cost. Consumers in such markets will be faced by prices that exceed marginal cost, and the allocation of resources will be inefficient.

An imperfectly competitive private market will produce less of a good than is efficient. As we saw in the chapter on monopoly, government agencies seek to prohibit monopoly in most markets and to regulate the prices charged by those monopolies that are permitted. Government policy toward monopoly is discussed more fully in a later chapter.

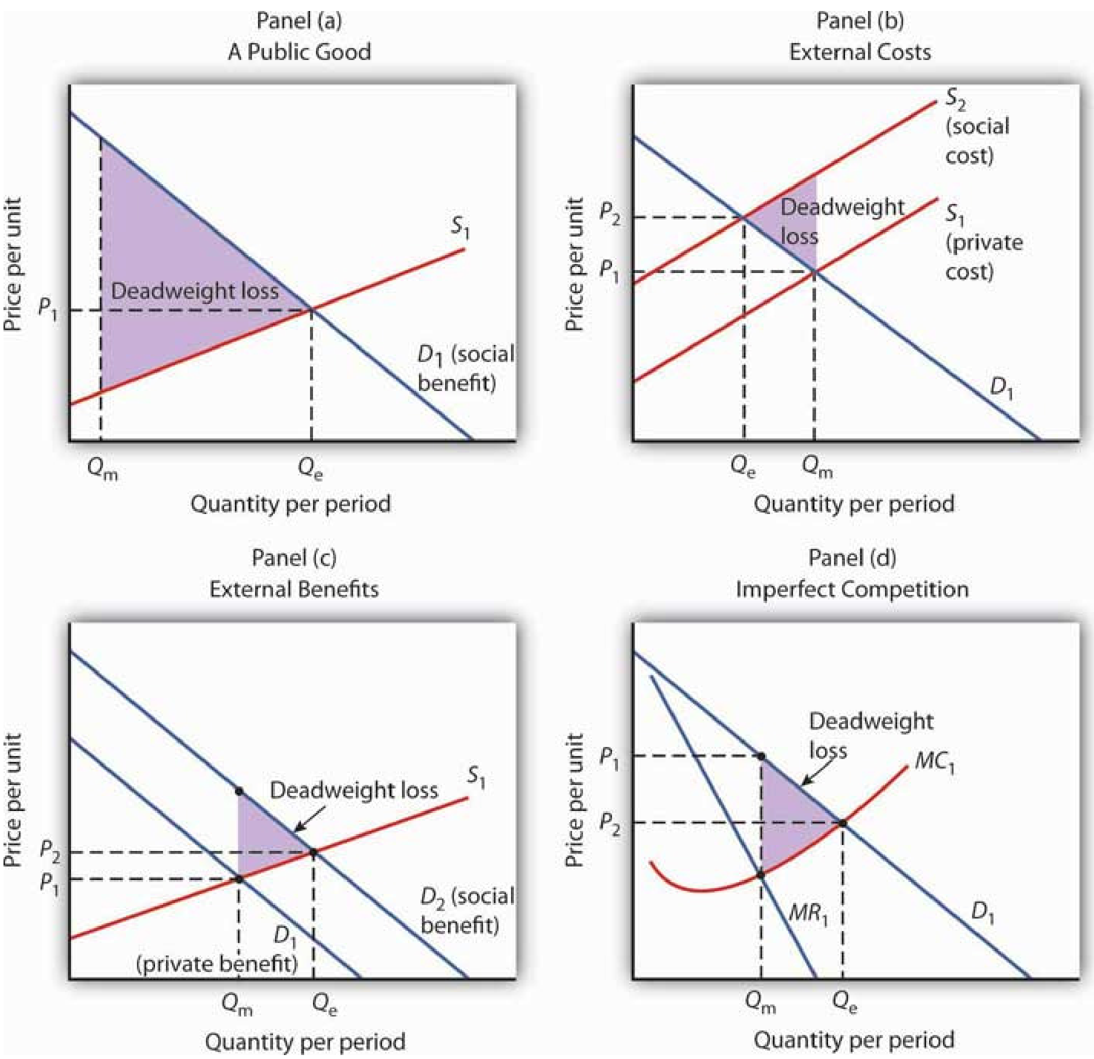
Assessing Government Responses to Market Failure

In each of the models of market failure we have reviewed here—public goods, external costs and benefits, and imperfect competition—the market may fail to achieve the efficient result. There is a potential for government intervention to move inefficient markets closer to the efficient solution.

Figure 15.3 reviews the potential gain from government intervention in cases of market failure. In each case, the potential gain is the deadweight loss resulting from market failure; government intervention may prevent or limit this deadweight loss. In each panel, the deadweight loss resulting from market failure is shown as a shaded triangle.

FIGURE 15.3 **Correcting Market Failure**

In each panel, the potential gain from government intervention to correct market failure is shown by the deadweight loss avoided, as given by the shaded triangle. In Panel (a), we assume that a private market produces Q_m units of a public good. The efficient level, Q_e, is defined by the intersection of the demand curve D_1 for the public good and the supply curve S_1. Panel (b) shows that if the production of a good generates an external cost, the supply curve S_1 reflects only the private cost of the good. The market will produce Q_m units of the good at price P_1. If the public sector finds a way to confront producers with the social cost of their production, then the supply curve shifts to S_2, and production falls to the efficient level Q_e. Notice that this intervention results in a higher price, P_2, which confronts consumers with the real cost of producing the good. Panel (c) shows the case of a good that generates external benefits. Purchasers of the good base their choices on the private benefit, and the market demand curve is D_1. The market quantity is Q_m. This is less than the efficient quantity, Q_e, which can be achieved if the activity that generates external benefits is subsidized. That would shift the market demand curve to D_2, which intersects the market supply curve at the efficient quantity. Finally, Panel (d) shows the case of a monopoly firm that produces Q_m units and charges a price P_1. The efficient level of output, Q_e, could be achieved by imposing a price ceiling at P_2. As is the case in each of the other panels, the potential gain from such a policy is the elimination of the deadweight loss shown as the shaded area in the exhibit.



Panel (a) of Figure 15.3 illustrates the case of a public good. The market will produce some of the public good; suppose it produces the quantity Q_m. But the demand curve that reflects the social benefits of the public good, D_1, intersects the supply curve at Q_e; that is the efficient quantity of the good. Public sector provision of a public good may move the quantity closer to the efficient level.

Panel (b) shows a good that generates external costs. Absent government intervention, these costs will not be reflected in the market solution. The supply curve, S_1, will be based only on the private costs associated with the good. The market will produce Q_m units of the good at a price P_1. If the government were to confront producers with the external cost of the good, perhaps with a tax on the activity that creates the cost, the supply curve would shift to S_2 and reflect the social cost of the good. The quantity would fall to the efficient level, Q_e, and the price would rise to P_2.

Panel (c) gives the case of a good that generates external benefits. The demand curve revealed in the market, D_1, reflects only the private benefits of the good. Incorporating the external benefits of the good gives us the demand curve D_2 that reflects the social benefit of the good. The market's output of Q_m units of the good falls short of the efficient level Q_e. The government may seek to move the market solution toward the efficient level through subsidies or other measures to encourage the activity that creates the external benefit.

Finally, Panel (d) shows the case of imperfect competition. A firm facing a downward-sloping demand curve such as D_1 will select the output Q_m at which the marginal cost curve MC_1 intersects the marginal revenue curve MR_1. The government may seek to move the solution closer to the efficient level, defined by the intersection of the marginal cost and demand curves.

While it is important to recognize the potential gains from government intervention to correct market failure, we must recognize the difficulties inherent in such efforts. Government officials may lack the information they need to select the efficient solution. Even if they have the information, they may have goals other than the efficient allocation of resources. Each instance of government intervention involves an interaction with utility-maximizing consumers and profit-maximizing firms, none of whom can be assumed to be passive participants in the process. So, while the potential exists for improved resource allocation in cases of market failure, government intervention may not always achieve it.

The late George Stigler, winner of the Nobel Prize for economics in 1982, once remarked that people who advocate government intervention to correct every case of market failure reminded him of the judge at an amateur singing contest who, upon hearing the first contestant, awarded first prize to the second. Stigler's point was that even though the market is often an inefficient allocator of resources, so is the government likely to be. Government may improve on what the market does; it can also make it worse. The choice between the market's allocation and an allocation with government intervention is always a choice between imperfect alternatives. We will examine the nature of public sector choices later in this chapter and explore an economic explanation of why government intervention may fail to move market solutions closer to their efficient levels.

1.2 Merit and Demerit Goods

In some cases, the public sector makes a determination that people should consume more of some goods and services and less of others, even in the absence of market failure. This is a normative judgment, one that presumes that consumers are not always the best judges of what is good, or bad, for them.

Merit goods are goods whose consumption the public sector promotes, based on a presumption that many individuals do not adequately weigh the benefits of the good and should thus be induced to consume more than they otherwise would. Many local governments support symphony concerts, for example, on grounds that the private market would not provide an adequate level of these cultural activities.

merit good

Goods whose consumption the public sector promotes, based on a presumption that many individuals do not adequately weigh the benefits of the good.

Indeed, government provision of some merit goods is difficult to explain. Why, for example, do many local governments provide tennis courts but not bowling alleys, golf courses but not auto racetracks, or symphony halls but not movie theaters? One possible explanation is that some consumers—those with a fondness for tennis, golf, and classical music—have been more successful than others in persuading their fellow citizens to assist in funding their preferred activities.

Demerit goods are goods whose consumption the public sector discourages, based on a presumption that individuals do not adequately weigh all the costs of these goods and thus should be induced to consume less than they otherwise would. The consumption of such goods may be prohibited, as in the case of illegal drugs, or taxed heavily, as in the case of cigarettes and alcohol.

demerit good

Goods whose consumption the public sector discourages, based on a presumption that individuals do not adequately weigh all the costs of these goods.

1.3 Income Redistribution

The proposition that a private market will allocate resources efficiently if the efficiency condition is met always comes with a qualification: the allocation of resources will be efficient *given the initial distribution of income*. If 5% of the people receive 95% of the income, it might be efficient to allocate roughly

95% of the goods and services produced to them. But many people (at least 95% of them!) might argue that such a distribution of income is undesirable and that the allocation of resources that emerges from it is undesirable as well.

There are several reasons to believe that the distribution of income generated by a private economy might not be satisfactory. For example, the incomes people earn are in part due to luck. Much income results from inherited wealth and thus depends on the family into which one happens to have been born. Likewise, talent is distributed in unequal measure. Many people suffer handicaps that limit their earning potential. Changes in demand and supply can produce huge changes in the values—and the incomes—the market assigns to particular skills. Given all this, many people argue that incomes should not be determined solely by the marketplace.

A more fundamental reason for concern about income distribution is that people care about the welfare of others. People with higher incomes often have a desire to help people with lower incomes. This preference is demonstrated in voluntary contributions to charity and in support of government programs to redistribute income.

A public goods argument can be made for government programs that redistribute income. Suppose that people of all income levels feel better off knowing that financial assistance is being provided to the poor and that they experience this sense of well-being whether or not they are the ones who provide the assistance. In this case, helping the poor is a public good. When the poor are better off, other people feel better off; this benefit is nonexclusive. One could thus argue that leaving private charity to the marketplace is inefficient and that the government should participate in income redistribution. Whatever the underlying basis for redistribution, it certainly occurs. The governments of every country in the world make some effort to redistribute income.

means-tested transfer payments

Transfer payment for which the recipient qualifies on the basis of income.

non-means-tested transfer payments

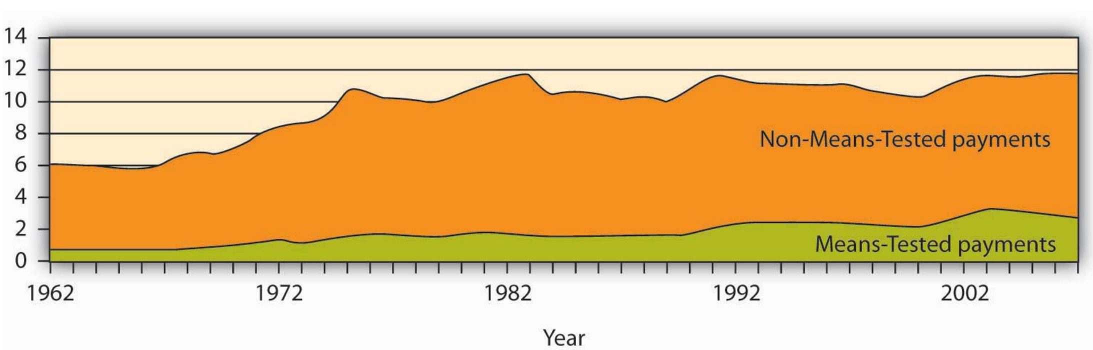
Transfer payment for which income is not a qualifying factor.

Programs to redistribute income can be divided into two categories. One transfers income to poor people; the other transfers income based on some other criterion. A **means-tested transfer payment** is one for which the recipient qualifies on the basis of income; means-tested programs transfer income from people who have more to people who have less. The largest means-tested program in the United States is Medicaid, which provides health care to the poor. Other means-tested programs include Temporary Assistance to Needy Families (TANF) and food stamps. A **non-means-tested transfer payment** is one for which income is not a qualifying factor. Social Security, a program that taxes workers and their employers and transfers this money to retired workers, is the largest non-means-tested transfer program. Indeed, it is the largest transfer program in the United States. It transfers income from working families to retired families. Given that retired families are, on average, wealthier than working families, Social Security is a somewhat regressive program. Other non-means tested transfer programs include Medicare, unemployment compensation, and programs that aid farmers.

Figure 15.4 shows federal spending on means-tested and non-means-tested programs as a percentage of GDP, the total value of output, since 1962. As the chart suggests, the bulk of income redistribution efforts in the United States are non-means-tested programs.

FIGURE 15.4 Federal Transfer Payment Spending

The chart shows federal means-tested and non-means-tested transfer payment spending as a percentage of GDP from 1962–2007.



Source: Congressional Budget Office, The Budget and Economic Outlook: Fiscal Years 2004–2013 (Jan., 2003), Table F-10p. 157; thereafter January, 2008, Table F-10 with means-tested as medicaid plus income security and non-means tested everything else.

The fact that most transfer payments in the United States are not means-tested leads to something of a paradox: some transfer payments involve taxing people whose incomes are relatively low to give to people whose incomes are relatively high. Social Security, for example, transfers income from people who are working to people who have retired. But many retired people enjoy higher incomes than working people in the United States. Aid to farmers, another form of non-means-tested payments, transfers income to farmers, who on average are wealthier than the rest of the population. These situations have come about because of policy decisions, which we discuss later in the chapter.

KEY TAKEAWAYS

- One role of government is to correct problems of market failure associated with public goods, external costs and benefits, and imperfect competition.
- Government intervention to correct market failure always has the potential to move markets closer to efficient solutions, and thus reduce deadweight losses. There is, however, no guarantee that these gains will be achieved.
- Governments may seek to alter the provision of certain goods and services based on a normative judgment that consumers will consume too much or too little of the goods. Goods for which such judgments are made are called merit or demerit goods.
- Governments redistribute income through transfer payments. Such redistribution often goes from people with higher incomes to people with lower incomes, but other transfer payments go to people who are relatively better off.

TRY IT!

Here is a list of actual and proposed government programs. Each is a response to one of the justifications for government activity described in the text: correction of market failure (due to public goods, external costs, external benefits, or imperfect competition), encouragement or discouragement of the consumption of merit or demerit goods, and redistribution of income. In each case, identify the source of demand for the activity described.

1. The Justice Department sought to prevent Microsoft Corporation from releasing Windows '98, arguing that the system's built-in internet browser represented an attempt by Microsoft to monopolize the market for browsers.
2. In 2004, Congress considered a measure that would extend taxation of cigarettes to vendors that sell cigarettes over the Internet.
3. The federal government engages in research to locate asteroids that might hit the earth, and studies how impacts from asteroids could be prevented.
4. The federal government increases spending for food stamps for people whose incomes fall below a certain level.
5. The federal government increases benefits for recipients of Social Security.
6. The Environmental Protection Agency sets new standards for limiting the emission of pollutants into the air.
7. A state utilities commission regulates the prices charged by utilities that provide natural gas to homes and businesses.

Case in Point: "Fixing" the Gasoline Market

Moderating the price of gasoline is not an obvious mission for the government in a market economy. But, in an economy in which angry voters wield considerable influence, trying to fix rising gasoline prices can turn into a task from which a wise politician does not shrink.

By the summer of 2008, crude oil was selling for more than $140 per barrel. Gasoline prices in the United States were flirting with the $4 mark. There were perfectly good market reasons for the run-up in prices. World oil demand has been rising each year, with China and India two of the primary sources of increased demand. The world's ability to produce oil is limited and tensions in the Middle East were also adding doubts about getting those supplies to market. Ability to produce gasoline is limited as well. The United States has not built a new oil refinery in more than 30 years.

But, when oil prices rise, economic explanations seldom carry much political clout. Predictably, the public demands a response from its political leaders—and gets it.

Largely Democratic Congressional proposals in 2008 included such ideas as: a bill to classify the Organization of Petroleum Exporting Countries (OPEC) as an illegal monopoly in violation of U.S. antitrust laws, taxing "excessive" profits of oil companies, investigating possible price gouging, and banning speculative trading in oil futures. With an overwhelming majority on both sides of the aisle, Congress passed a bill to suspend adding oil to the Strategic Petroleum Reserve—a 727 million gallon underground reserve designed for use in national emergencies. President Bush in 2008 was against this move, though in 2006, when gas prices were approaching $3 a gallon, he supported a similar move. Whether or not to offer a "tax holiday" on the 18.4 cents per gallon federal gas tax stymied some politicians during the 2008 presidential campaign because Hillary Clinton, a Democrat, and John McCain, a Republican, supported it, while Barack Obama, a Democrat, was against it. Mostly Republican proposals to allow offshore drilling and exploration in the Arctic National Wildlife Refuge also received attention.

These measures were unlikely to have much affect on gas prices, especially in the short-term. For example, the federal government would normally in a two-month period deposit 10 million gallons of gasoline in the strategic reserve; consumption in the United States is about 20 million gallons of gasoline per day. World gasoline consumption is about 87 million gallons per day. Putting an additional 10 million gallons into a global market which will consume about 5 billion gallons in a 60-day period is not likely to have any measurable impact.

The higher oil prices were very good for oil companies. Exxon Mobil, the largest publicly traded oil company in the United States, reported profits of nearly $11 billion for the first quarter of 2008. Whenever oil prices rise sharply, there are always cries of "price gouging." But, repeated federal investigations of the industry have failed to produce any evidence that such gouging has occurred.

Meanwhile, market forces responding to the higher gasoline prices are already at work. Gasoline producers are looking at cellulosic ethanol, which can be produced from materials such as wood chips, corn stalks, and rice straw. Automobile producers are examining "plug-in" hybrids—cars whose batteries could be charged not just by driving but by plugging the car in a garage. The goal is to have a car that could go some distance on its battery before starting to use any gasoline. Consumers are doing their part. Gasoline consumption in the United States fell more than 4% by the summer of 2008 from its level one year earlier.

These potential market responses are the sort of thing one would expect from rising fuel prices. Ultimately, it is difficult to see why gasoline prices should be a matter for public sector intervention. But, the public sector consists of people, and when those people become angry, the urge for intervention can become unstoppable.

Sources: Paul Davidson and Chris Woodyard, "Proposals To Cut Gas Prices Scrutinized," USA Today, May 11, 2006, p. 5B; Joseph Curl, "Bush Orders Suspension Of Gas Rules; Federal Probe To Look At Price-Gouging Charges," The Washington Times, April 26, 2007, p. A1; David M. Herszenhorn, "As Gasoline Prices Soar, Politicians Fall Back on Familiar Solutions," The New York Times, May 3, 2008, p. A16; Richard Simon, "The Nation; Mixing Oil and Politics; Congress Votes To Stop Shipments to the Nation's Reserve. The Move Could Save Motorists Some Money," Los Angeles Times, May 14, 2008, p. A18.

ANSWERS TO TRY IT! PROBLEMS

1. This is an attempt to deal with monopoly, so it is a response to imperfect competition.
2. Cigarettes are treated as a demerit good.
3. Protecting the earth from such a calamity is an example of a public good.
4. Food Stamps are a means-tested program to redistribute income.
5. Social Security is an example of a non-means-tested income redistribution program.
6. This is a response to external costs.
7. This is a response to monopoly, so it falls under the imperfect competition heading.

2. FINANCING GOVERNMENT

LEARNING OBJECTIVES

1. **Explain the ability-to-pay and the benefits-received principles of taxation.**
2. **Distinguish among regressive, proportional, and progressive taxes.**
3. **Define tax incidence analysis and explain and illustrate the conditions under which the burden of an excise tax falls mainly on buyers or sellers.**

If government services are to be provided, people must pay for them. The primary source of government revenue is taxes. In this section we examine the principles of taxation, compare alternative types of taxes, and consider the question of who actually bears the burden of taxes.

In addition to imposing taxes, governments obtain revenue by charging **user fees**, which are fees levied on consumers of government-provided services. The tuition and other fees charged by public universities and colleges are user fees, as are entrance fees at national parks. Finally, government agencies might obtain revenue by selling assets or by holding bonds on which they earn interest.

user fees

Fees levied on consumers of government-provided services.

2.1 Principles of Taxation

Virtually anything can be taxed, but what should be taxed? Are there principles to guide us in choosing a system of taxes?

Jean-Baptiste Colbert, a minister of finance in seventeenth-century France, is generally credited with one of the most famous principles of taxation:

> *"The art of taxation consists in so plucking the goose as to obtain the largest possible amount of feathers with the smallest possible amount of hissing."*

Economists, who do not typically deal with geese, cite two criteria for designing a tax system. The first is based on the ability of people to pay taxes and the second focuses on the benefits they receive from particular government services.

Ability to Pay

ability-to-pay principle

Principle that holds that people with more income should pay more taxes.

The **ability-to-pay principle** holds that people with more income should pay more taxes. As income rises, the doctrine asserts, people are able to pay more for public services; a tax system should therefore be constructed so that taxes rise too. Wealth, the total of assets less liabilities, is sometimes used as well as income as a measure of ability to pay.

The ability-to-pay doctrine lies at the heart of tax systems that link taxes paid to income received. The relationship between taxes and income may take one of three forms: taxes can be regressive, proportional, or progressive.

Regressive Tax

regressive tax

A tax that takes a higher percentage of income as income falls.

A **regressive tax** is one that takes a higher percentage of income as income falls. Taxes on cigarettes, for example, are regressive. Cigarettes are an inferior good—their consumption falls as incomes rise. Thus, people with lower incomes spend more on cigarettes than do people with higher incomes. The cigarette taxes paid by low-income people represent a larger share of their income than do the cigarette taxes paid by high-income people and are thus regressive.

Proportional Tax

proportional tax

Tax that takes a fixed percentage of income, no matter what the level of income.

A **proportional tax** is one that takes a fixed percentage of income. Total taxes rise as income rises, but taxes are equal to the same percentage no matter what the level of income. Some people argue that the U.S. income tax system should be changed into a flat tax system, a tax that would take the same percentage of income from all taxpayers. Such a tax would be a proportional tax.

Progressive Tax

progressive tax

A tax that takes a higher percentage of income as income rises.

A **progressive tax** is one that takes a higher percentage of income as income rises. The federal income tax is an example of a progressive tax. Table 15.1 shows federal income tax rates for various brackets of income for a family of four in 2007. Such a family paid no income tax at all if its income fell below $24,300. At higher income levels, families faced a higher percentage tax rate. Any income over $374,000, for example, was taxed at a rate of 35%. Whether or not to make the tax system more progressive was a major debating point during the U.S. presidential election of 2008.

TABLE 15.1 Federal Income Tax Brackets, 2007

The federal income tax is progressive. The percentage tax rate rises as adjusted gross income rises.

2007 adjusted gross income (family of four)	Personal income tax rate applied to bracket
Less than $24,300	Zero (family may receive earned income credit)
$24,300–$88,000	15%
$88,000–152,800	25%
$152,800–$220,150	28%
$220,150–$374,000	33%
Greater than $374,000	35%

While a pure flat tax would be proportional, most proposals for such a tax would exempt some income from taxation. Suppose, for example, that households paid a "flat" tax of 20% on all income over $40,000 per year. This tax would be progressive. A household with an income of $25,000 per year would pay no tax. One with an income of $50,000 per year would pay a tax of $2,000 (.2 times $10,000), or 4% of its income. A household with an income of $100,000 per year would pay a tax of $12,000 (.2 times $60,000) per year, or 12% of its income. A flat tax with an income exemption would thus be a progressive tax.

Benefits Received

benefits-received principle

Principle that holds that a tax should be based on the benefits received from the government services funded by the tax.

An alternative criterion for establishing a tax structure is the **benefits-received principle**, which holds that a tax should be based on the benefits received from the government services funded by the tax.

Local governments rely heavily on taxes on property, in large part because the benefits of many local services, including schools, streets, and the provision of drainage for wastewater, are reflected in higher property values. Suppose, for example, that public schools in a particular area are especially good. People are willing to pay more for houses served by those schools, so property values are higher; property owners benefit from better schools. The greater their benefit, the greater the property tax they pay. The property tax can thus be viewed as a tax on benefits received from some local services.

User fees for government services apply the benefits-received principle directly. A student paying tuition, a visitor paying an entrance fee at a national park, and a motorist paying a highway toll are all paying to consume a publicly provided service; they are thus paying directly for something from which they expect to benefit. Such fees can be used only for goods for which exclusion is possible; a user fee could not be applied to a service such as national defense.

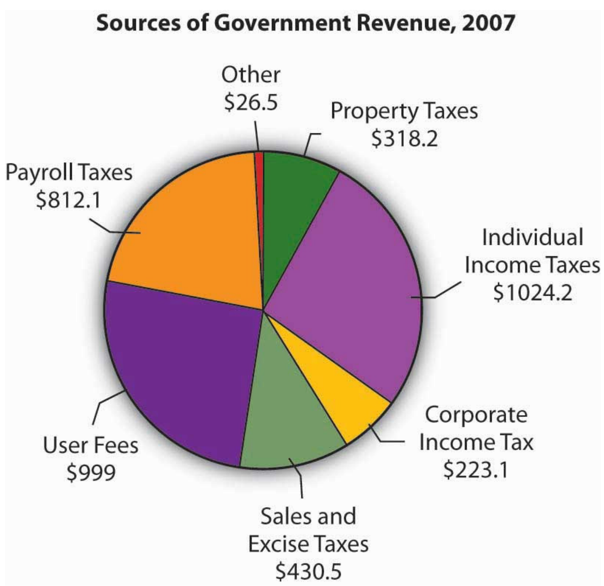
Income taxes to finance public goods may satisfy both the ability-to-pay and benefits-received principles. The demand for public goods generally rises with income. Thus, people with higher incomes benefit more from public goods. The benefits-received principle thus suggests that taxes should rise with income, just as the ability-to-pay principle does. Consider, for example, an effort financed through income taxes by the federal government to clean up the environment. People with higher incomes will pay more for the cleanup than people with lower incomes, consistent with the ability-to-pay principle. Studies by economists consistently show that people with higher incomes have a greater demand for environmental improvement than do people with lower incomes—a clean environment is a normal good. Requiring people with higher incomes to pay more for the cleanup can thus be justified on the benefits-received principle as well.

Certainly taxes cannot respond precisely to benefits received. Neither the ability-to-pay nor the benefits-received doctrine gives us a recipe for determining just what each person "should" pay in taxes, but these doctrines give us a framework for thinking about the justification for particular taxes.

2.2 Types of Taxes

FIGURE 15.6 **Sources of Government Revenue, 2007**

The chart shows sources of revenue for federal, state, and local governments in the United States. The data omit revenues from government-owned utilities and liquor stores. All figures are in billions of dollars.



Source: U.S. Bureau of the Census, Statistical Abstract of US, 2008 (online) Tables 422 and 461.

income tax

Taxes imposed on the income earned by a person or firm.

property tax

Taxes imposed on assets.

sales tax

Taxes imposed on the value of goods sold.

excise tax

Taxes imposed on specific goods or services.

taxable income

The amount of income that is actually subject to any tax.

marginal tax rate

The tax rate that would apply to an additional $1 of taxable income earned.

It is hard to imagine anything that has not been taxed at one time or another. Windows, closets, buttons, junk food, salt, death—all have been singled out for special taxes. In general, taxes fall into one of four primary categories. **Income taxes** are imposed on the income earned by a person or firm; **property taxes** are imposed on assets; **sales taxes** are imposed on the value of goods sold; and **excise taxes** are imposed on specific goods or services. Figure 15.6 shows the major types of taxes financing all levels of government in the United States.

Personal Income Taxes

The federal personal income tax is the largest single source of tax revenue in the United States; most states and many cities tax income as well. All income tax systems apply a variety of exclusions to a taxpayer's total income before arriving at **taxable income**, the amount of income that is actually subject to the tax. In the U.S. federal income tax system, for example, a family deducted $3,200 from total income earned in 2005 for each member of the family as part of its computation of taxable income.

Income taxes can be structured to be regressive, proportional, or progressive. Income tax systems in use today are progressive.

In analyzing the impact of a progressive tax system on taxpayer choice, economists focus on the **marginal tax rate**. This is the tax rate that would apply to an additional $1 of taxable income earned. Suppose an individual was earning taxable income of $8,025 and paid federal income taxes of $802.50, or 10% of taxable income (we are ignoring exemptions that would eliminate taxes for such an individual). If the taxpayer were to receive $100 more of taxable income, however, that $100 would be taxed at a rate of 15%, the rate that applied in 2008 to taxable incomes between $8,025–$32,550 for individuals. That person thus faced a marginal tax rate of 15%.

Economists argue that choices are made at the margin; it is thus the marginal tax rate that is most likely to affect decisions. Say that the individual in our example is considering taking on additional work that would increase his or her income to $15,025 per year. With a marginal tax rate of 15%, the individual would keep $5,950 of the additional $7,000 earned. It is that $5,950 that the individual will weigh against the opportunity cost in forgone leisure in deciding whether to do the extra work.

Property Taxes

Property taxes are taxes imposed on assets. Local governments, for example, generally impose a property tax on business and personal property. A government official (typically a local assessor) determines the property's value, and a proportional tax rate is then applied to that value.

Property ownership tends to be concentrated among higher income groups; economists generally view property taxes as progressive. That conclusion, however, rests on assumptions about who actually pays the tax, an issue examined later in this chapter.

Sales Taxes

Sales taxes are taxes imposed as a percentage of firms' sales and are generally imposed on retail sales. Some items, such as food and medicine, are often exempted from sales taxation.

People with lower incomes generally devote a larger share of their incomes to consumption of goods covered by sales taxes than do people with higher incomes. Sales taxes are thus likely to be regressive.

Excise Taxes

An excise tax is imposed on specific items. In some cases, excise taxes are justified as a way of discouraging the consumption of demerit goods, such as cigarettes and alcoholic beverages. In other cases, an excise tax is a kind of benefits-received tax. Excise taxes on gasoline, for example, are typically earmarked for use in building and maintaining highways, so that those who pay the tax are the ones who benefit from the service provided.

The most important excise tax in the United States is the payroll tax imposed on workers' earnings. In 2007, the payroll tax was 12.4% and was levied on incomes up to $97,500. The Medicare portion of the payroll tax, 2.9%, was levied on all earned wages without limit. Half of the payroll tax is charged to employers, half to employees. The proceeds of this excise on payrolls finance Social Security and Medicare benefits. Almost two-thirds of U. S. households pay more in payroll taxes than in any other taxes.

2.3 Tax Incidence Analysis

Next time you purchase an item at a store, notice the sales tax imposed by your state, county, and city. The clerk rings up the total, then adds up the tax. The store is the entity that "pays" the sales tax, in the sense that it sends the money to the government agencies that imposed it, but you are the one who actually foots the bill—or are you? Is it possible that the sales tax affects the price of the item itself?

These questions relate to **tax incidence analysis**, a type of economic analysis that seeks to determine where the actual burden of a tax rests. Does the burden fall on consumers, workers, owners of capital, owners of natural resources, or owners of other assets in the economy? When a tax imposed on a good or service increases the price by the amount of the tax, the burden of the tax falls on consumers. If instead it lowers wages or lowers prices for some of the other factors of production used in the production of the good or service taxed, the burden of the tax falls on owners of these factors. If the tax does not change the product's price or factor prices, the burden falls on the owner of the firm—the owner of capital. If prices adjust by a fraction of the tax, the burden is shared.

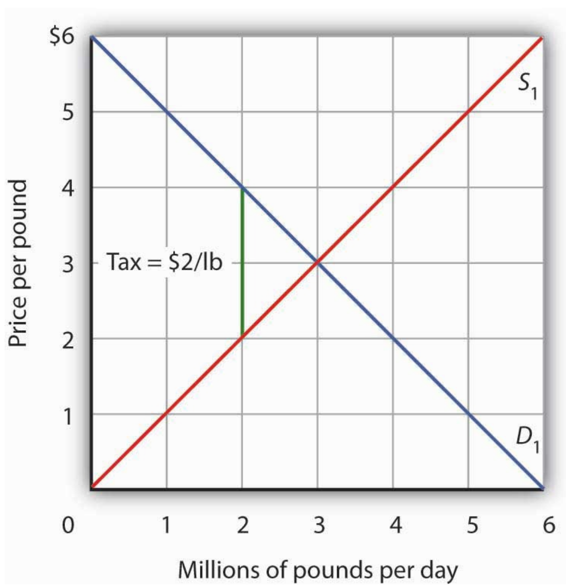
tax incidence analysis

A type of economic analysis that seeks to determine where the actual burden of a tax rests.

Figure 15.7 gives an example of tax incidence analysis. Suppose D_1 and S_1 are the demand and supply curves for beef. The equilibrium price is $3 per pound; the equilibrium quantity is 3 million pounds of beef per day. Now suppose an excise tax of $2 per pound of beef is imposed. It does not matter whether the tax is levied on buyers or on sellers of beef; the important thing to see is that the tax drives a $2 per pound "wedge" between the price buyers pay and the price sellers receive. This tax is shown as the vertical green line in the exhibit; its height is $2.

FIGURE 15.7 Tax Incidence in the Model of Demand and Supply

Suppose the market price of beef is $3 per pound; the equilibrium quantity is 3 million pounds per day. Now suppose an excise tax of $2 per pound is imposed, shown by the vertical green line. We insert this tax wedge between the demand and supply curves. It raises the market price to $4 per pound, suggesting that buyers pay half the tax in the form of a higher price. Sellers receive a price of $2 per pound; they pay half the tax by receiving a lower price. The equilibrium quantity falls to 2 million pounds per day.

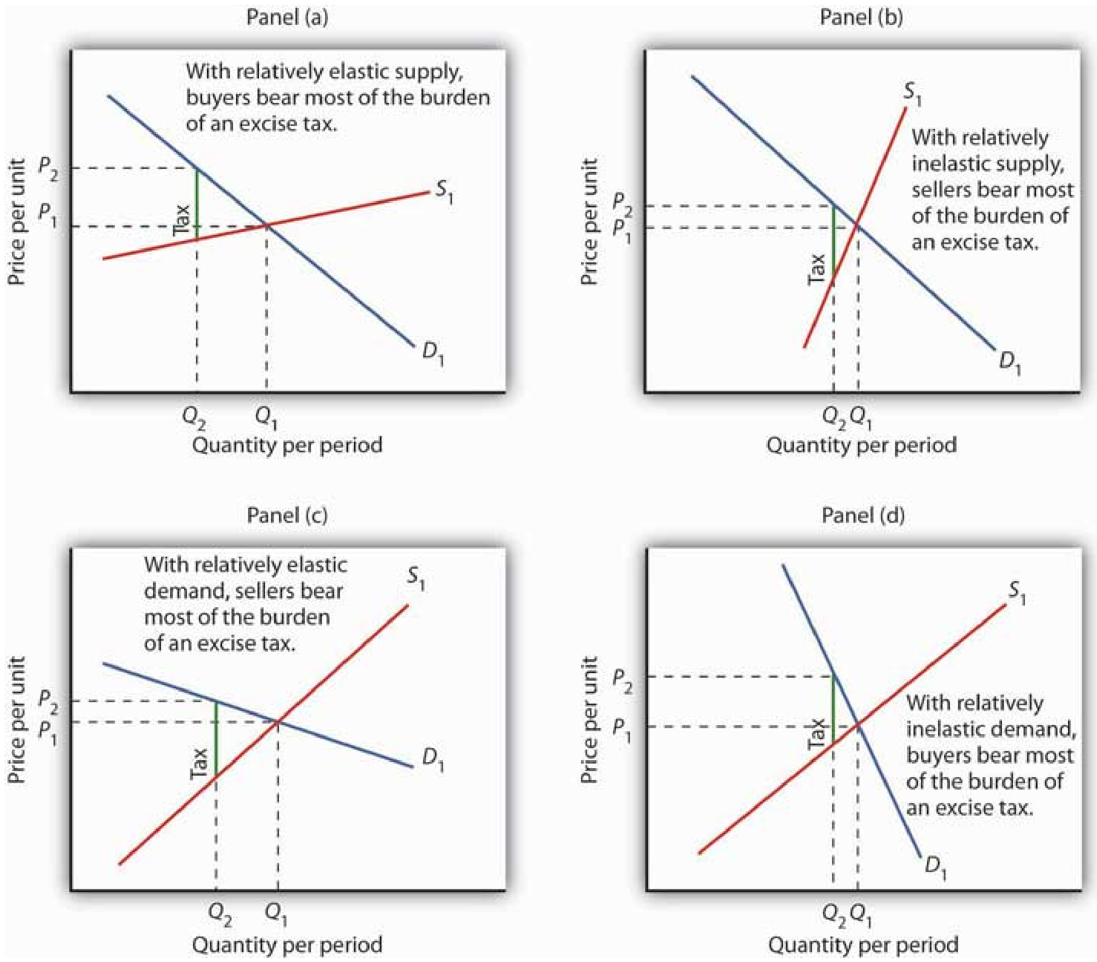


We insert our tax "wedge" between the demand and supply curves. In our example, the price paid by buyers rises to $4 per pound. The price received by sellers falls to $2 per pound; the other $2 goes to the government. The quantity of beef demanded and supplied falls to 2 million pounds per day. In this case, we conclude that buyers bear half the burden of the tax (the price they pay rises by $1 per pound), and sellers bear the other half (the price they receive falls by $1 per pound). In addition to the change in price, a further burden of the tax results from the reduction in consumer and in producer surplus. We have not shown this reduction in the graph.

Figure 15.8 shows how tax incidence varies with the relative elasticities of demand and supply. All four panels show markets with the same initial price, P_1, determined by the intersection of demand D_1 and supply S_1. We impose an excise tax, given by the vertical green line. As before, we insert this tax wedge between the demand and supply curves. We assume the amount of the tax per unit is the same in each of the four markets.

FIGURE 15.8 Tax Incidence and the Elasticity of Demand and of Supply

We show the effect of an excise tax, given by the vertical green line, in the same way that we did in Figure 15.7. We see that buyers bear most of the burden of such a tax in cases of relatively elastic supply (Panel (a)) and of relatively inelastic demand (Panel (d)). Sellers bear most of the burden in cases of relatively inelastic supply (Panel (b)) and of relatively elastic demand (Panel (c)).



In Panel (a), we have a market with a relatively elastic supply curve S_1. When we insert our tax wedge, the price rises to P_2; the price increase is nearly as great as the amount of the tax. In Panel (b), we have the same demand curve as in Panel (a), but with a relatively inelastic supply curve S_2. This time the price paid by buyers barely rises; sellers bear most of the burden of the tax. When the supply curve is relatively elastic, the bulk of the tax burden is borne by buyers. When supply is relatively inelastic, the bulk of the burden is borne by sellers.

Panels (c) and (d) of the exhibit show the same tax imposed in markets with identical supply curves S_1. With a relatively elastic demand curve D_1 in Panel (c) (notice that we are in the upper half, that is, the elastic portion of the curve), most of the tax burden is borne by sellers. With a relatively inelastic demand curve D_1 in Panel (d) (notice that we are in the lower half, that is, the inelastic portion of the curve), most of the burden is borne by buyers. If demand is relatively elastic, then sellers bear more of the burden of the tax. If demand is relatively inelastic, then buyers bear more of the burden.

The Congressional Budget Office (CBO) has prepared detailed studies of the federal tax system. Using the tax laws in effect in August 2004, it ranked the U.S. population according to income and then divided the population into quintiles (groups containing 20% of the population). Then, given the federal tax burden imposed by individual income taxes, payroll taxes for social insurance, corporate income taxes, and excise taxes on each quintile and the income earned by people in that quintile, it projected the average tax rate facing that group in 2006. The study assigned taxes on the basis of who bears the burden, not on who pays the tax. For example, many studies argue that, even though businesses pay half of the payroll taxes, the burden of payroll taxes actually falls on households. The reason is that the supply curve of labor is relatively inelastic, as shown in Panel (b) of Figure 15.8. Taking these adjustments into account, the CBO's results, showing progressivity in federal taxes, are reported in Table 15.2.

TABLE 15.2 Federal Tax Burdens in the United States

In a regressive tax system, people in the lowest quintiles face the highest tax rates. A proportional system imposes the same rates on everyone; a progressive system imposes higher rates on people in higher deciles. The table gives estimates by the CBO of the burden on each quintile of federal taxes in 2006. As you can see, the tax structure in the United States is progressive.

Income category	Households (number, millions)	Average pretax comprehensive household income	Effective federal tax rate, 2006 (percent)
Lowest quintile	24.0	$18,568	5.6
Second quintile	22.8	$42,619	12.1
Middle quintile	23.3	$64,178	15.7
Fourth quintile	23.2	$94,211	19.8
Highest quintile	24.3	$227,677	26.5
All quintiles	118.3	$89,476	21.6

Source: CBO, Effective Federal Tax Rates under Current Law, 2001 to 2014, August, 2004, Table 2 and Table A-1 (adjusted by authors using CBO assumptions concerning rates of growth of income and households). Numbers of households do not add up to total because of excluded categories. Quintiles contains equal numbers of people.

KEY TAKEAWAYS

- The primary principles of taxation are the ability-to-pay and benefits-received principles.
- The percentage of income taken by a regressive tax rises as income falls. A proportional tax takes a constant percentage of income regardless of income level. A progressive tax takes a higher percentage of income as taxes as incomes rise.
- The marginal tax rate is the tax rate that applies to an additional dollar of income earned.
- Tax incidence analysis seeks to determine who ultimately bears the burden of a tax.
- The major types of taxes are income taxes, sales taxes, property taxes, and excise taxes.
- Buyers bear most of the burden of an excise tax when supply is relatively elastic and when demand is relatively inelastic; sellers bear most of the burden when supply is relatively inelastic and when demand is relatively elastic.
- The federal tax system in the United States is progressive.

TRY IT!

Consider three goods, A, B, and C. The prices of all three goods are determined by demand and supply (that is, the three industries are perfectly competitive) and equal $100. The supply curve for good A is perfectly elastic; the supply curve for good B is a typical, upward-sloping curve; and the supply curve for good C is perfectly inelastic. Suppose the federal government imposes a tax of $20 per unit on suppliers of each good. Explain and illustrate graphically how the tax will affect the price of each good in the short run. Show whether the equilibrium quantity will rise, fall, or remain unchanged. Who bears the burden of the tax on each good in the short run? (Hint: Review the chapter on the elasticity for a discussion of perfectly elastic and perfectly inelastic supply curves; remember that the tax increases variable cost by $20 per unit.)

Case in Point: What Are Marginal Tax Rates?

© 2010 Jupiterimages Corporation

We speak often of the importance of tax rates at the margin—of how much of an extra dollar earned through labor or interest on saving will be kept by the decision-maker. It turns out, however, that figuring out just what that marginal tax rate is is not an easy task.

Consider the difficulty of untangling just what those marginal tax rates are. First, Americans face a bewildering complex of taxes. They all face the federal income tax. Each state—and many cities—levy additional taxes on income. Then there is the FICA payroll tax, federal and state corporate income taxes, and excise taxes, as well as federal, state, and local sales taxes. A person trying to figure out his or her marginal tax rate cannot stop there. Gaining an additional dollar of income will affect not only taxes but eligibility for various transfer payment programs in the level of payments the individual or household can expect to receive. Given the enormous complexity involved, it is safe to say that no one really knows what his or her marginal rate is.

Economists Laurence J. Kotlikoff and David Rapson of Boston University have taken on the task of sorting out marginal tax rates for the United States. They used a commercial tax analysis program, Economic Security Planner™, and added their own computer programs to incorporate the effect of additional income on various transfer payment programs. Their analysis assumed the taxpayer lived in Massachusetts, but the general tenor of their results applies to people throughout the United States.

Consider a 60-year-old couple earning $10,000 per year. That couple is eligible for a variety of welfare programs. With food stamps, there is a dollar-for-dollar reduction in aid for each additional dollar of income earned. In effect, the couple faces an effective marginal tax rate of 100%. Considering all other taxes and welfare programs, the economists concluded that the couple faced a marginal tax rate of about 50% on labor income. Overall, they found that a pattern of marginal rates for various ages and income levels could be described in a single word: "bizarre."

The tables below give the economists' estimates of marginal rates for current year labor supply for a single individual and for couples with children at various incomes and ages. While the overall structure of taxes in the United States is progressive, the special treatment of welfare programs can add a strong element of regressivity.

Marginal Net Tax Rates on Current-Year Labor Supply (Couples, percentages)					
	Total Annual Household Earnings (000s)				
Age	***10***	***20***	***30***	***50***	***75***
30	−14.2	42.5	42.3	24.4	36.9
45	−11.4	41.7	41.8	35.8	36.1
60	50.9	32.0	36.3	36.3	45.5
Age	***100***	***150***	***200***	***300***	***500***
30	37.0	45.9	36.8	43.9	44.0
45	36.1	45.1	35.9	40.0	43.2
60	45.5	47.7	43.2	45.8	45.0

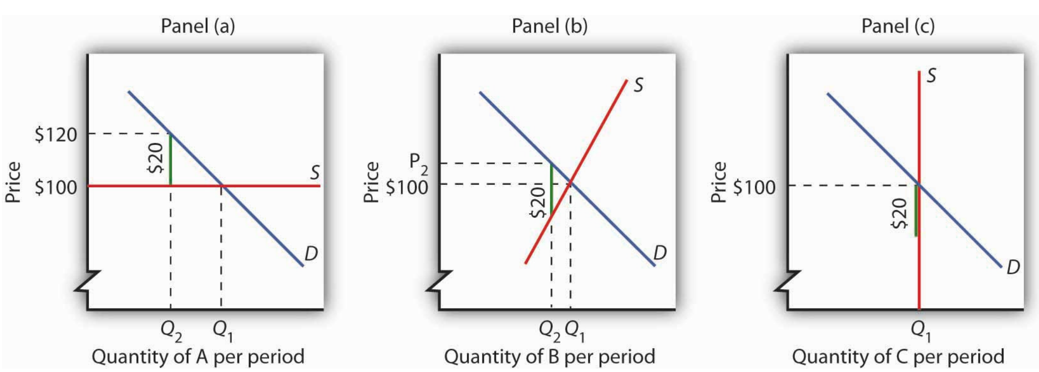
Source: Laurence J. Kotlikoff and David Rapson, "Does It Pay, At the Margin, to Work and Save?" NBER Tax Policy & the Economy, 2007, 21(1): 83–143. The tables shown here are Tables 4.2 and 4.3 in the article.

Marginal Net Tax Rates on Current-Year Labor Supply (Individuals, percentages)					
	Total Annual Household Earnings (000s)				
Age	***10***	***20***	***30***	***50***	***75***
30	72.3	42.9	42.9	37.0	37.0
45	−0.8	42.9	42.6	37.0	36.1
60	39.5	37.3	37.7	46.4	45.5
Age	***125***	***150***	***200***	***250***	
30	36.2	36.9	42.0	41.5	
45	36.1	36.5	42.0	41.5	
60	38.8	44.0	45.0	44.0	

Look again at our 60-year-old couple. It faces a very high marginal tax rate. A younger couple with the same income actually faces a negative marginal tax rate—increasing its labor income by a dollar actually increases its after-tax income by more than a dollar. Why the difference? The economists assumed that the younger couple would have children and thus qualify for a variety of programs, including the Earned Income Tax Credit. The couple at age 60 still faces the dollar-for-dollar reduction in payments in the Food Stamp program. No one designed these marginal incentives. They simply emerge from the bewildering mix of welfare and tax programs households face.

ANSWER TO TRY IT! PROBLEM

The tax adds a $20 wedge between the price paid by buyers and received by sellers. In Panel (a), the price rises to $120; the entire burden is borne by buyers. In Panel (c), the price remains $100; sellers receive just $80. Therefore, sellers bear the burden of the tax. In Panel (b), the price rises by less than $20, and the burden is shared by buyers and sellers. The relative elasticities of demand and supply determine whether the tax is borne primarily by buyers or sellers, or shared equally by both groups.



3. CHOICES IN THE PUBLIC SECTOR

LEARNING OBJECTIVES

1. **Compare public interest theory and public choice theory.**
2. **Use public choice theory to explain rational abstention and why legislative choices may serve special interests.**

public interest theory

Theory that assumes that the goal of government is to seek an efficient allocation of resources.

public choice theory

Body of economic thought based on the assumption that individuals involved in public sector choices make those choices to maximize their own utility.

rent-seeking behavior

The effort to influence public choices to advance one's own self-interest.

cost-benefit analysis

A type of economic analysis that seeks to quantify the costs and benefits of an activity.

How are choices made in the public sector? This section examines two perspectives on public sector choice. The first is driven by our examination of market failure. Choices in the public sector are a matter of locating problems of market failure, determining the efficient solution, and finding ways to achieve it. This approach, called the **public interest theory** of government, assumes that the goal of government is to seek an efficient allocation of resources.

An alternative approach treats public sector choices like private sector choices. The body of economic thought based on the assumption that individuals involved in public sector choices make those choices to maximize their own utility is called **public choice theory**. Public choice theory argues that individuals in the public sector make choices that maximize their utility—whether as voters, politicians, or bureaucrats, people seek solutions consistent with their self-interest. People who operate business firms may try to influence public sector choices to increase the profits of their firms. The effort to influence public choices to advance one's own self-interest is called **rent-seeking behavior**.

3.1 Public Interest Theory

In the approach to the analysis of public sector choices known as **public interest theory**, decision making is a technical matter. The task of government officials is to locate the efficient solution and find a way to move the economy to that point.

For a public good, the efficient solution occurs where the demand curve that reflects social benefits intersects the supply curve for producing the good; that is, the solution at quantity Q_e and price P_1 given in Panel (a) of Figure 15.3 Because this demand curve for a public good is not revealed in the market, the task for government officials is to find a way to estimate these curves and then to arrange for the production of the optimum quantity. For this purpose, economists have developed an approach called **cost-benefit analysis**, which seeks to quantify the costs and benefits of an activity. Public officials can use cost-benefit analysis to try to locate the efficient solution. In general, the efficient solution occurs where the net benefit of the activity is maximized.

Public sector intervention to correct market failure presumes that market prices do not reflect the benefits and costs of a particular activity. If those prices are generated by a market that we can regard as perfectly competitive, then the failure of prices to convey information about costs or benefits suggests that there is a free-rider problem on the demand side or an external cost problem on the supply side. In either case, it is necessary to estimate costs or benefits that are not revealed in the marketplace.

The public interest perspective suggests an approach in which policy makers identify instances of potential market failure and then look for ways to correct them. Public choice theory instead looks at what motivates the people making those policy choices.

3.2 The Public Choice Perspective

Public choice theory discards the notion that people in the public sector seek to maximize net benefits to society as a whole. Rather, it assumes that each participant in the public sector seeks to maximize his or her own utility. This section introduces the flavor of the public choice approach by examining two of its more important conclusions: that many people will abstain from voting, and that legislative choices are likely to serve special interests.

Economics and Voting: The Rational Abstention Problem

Public choice theory argues that individuals do not leave their self-interests behind when they enter the voting booth—or even when they are thinking about whether to go to the voting booth. The assumption of utility maximization by voters helps us to understand why most people do not vote in most elections.

Suppose your state is about to hold a referendum on expanded support for state recreation areas, to be financed by an increase in the state sales tax. Given your own likely use of these areas and the way in which you expect to be affected by the tax, you estimate that you will be better off if the program passes. In fact, you have calculated that the present value of your net benefits from the program is $1,000. Will you vote?

As a utility maximizer, you will vote if the marginal benefits to you of voting exceed the marginal costs. One benefit of voting is the possibility that your vote will cause the measure to be passed. That would be worth $1,000 to you. But $1,000 is a benefit to you of voting only if it is your vote that determines the outcome.

The probability that any statewide election will be decided by a single vote is, effectively, zero. State elections that are decided by as many as a few hundred votes are likely to be subject to several recounts, each of which is likely to produce a different result. The outcomes of extremely close elections are ordinarily decided in the courts or in legislative bodies; there is no chance that one vote would, in fact, determine the outcome. Thus, the $1,000 benefit that you expect to receive will not be a factor in your decision about whether to vote. The other likely benefit of voting is the satisfaction you receive from performing your duty as a citizen in a free society. There may be additional personal benefits as well from the chance to visit with other people in your precinct. The opportunity cost of voting would be the value of the best alternative use of your time, together with possible transportation costs.

The fact that no one vote is likely to determine the outcome means that a decision about whether to vote is likely to rest on individual assessments of the satisfactions versus the costs of voting. Most people making such decisions find the costs are greater. In most elections, most people who are eligible to vote do not vote. Public choice analysis suggests that such a choice is rational; a decision not to vote because the marginal costs outweigh the marginal benefits is called **rational abstention**.

rational abstention

A decision not to vote because the marginal costs outweigh the marginal benefits.

Rational abstention suggests there is a public sector problem of external benefits. Elections are a way of assessing voter preferences regarding alternative outcomes. An election is likely to do a better job of reflecting voter preferences when more people vote. But the benefits of an outcome that reflects the preferences of the electorate do not accrue directly to any one voter; a voter faces only some of the benefits of voting and essentially all of the costs. Voter turnouts are thus likely to be lower than is economically efficient.

In the 2000 presidential election, for example, just 50.7% of the voting-age population actually cast votes. President Bush received 47.9% of the vote, which means he was elected with the support of just 24% of the electorate. Mr. Bush actually received fewer votes than his opponent, Albert Gore, Jr. Mr. Bush, however, won a majority in the Electoral College. The Case in Point essay describes the 2000 election in more detail. Voter turnout was higher in the 2004 and 2008 presidential elections.

Legislative Choice and Special Interests

One alternative to having the general public vote on issues is to elect representatives who will make choices on their behalf. Public choice theory suggests that there are some difficulties with this option as well.

Suppose legislators seek to maximize the probability that they will be reelected. That requires that a legislator appeal to a majority of voters in his or her district. Suppose that each legislator can, at zero cost, learn the preferences of every voter in his or her district. Further, suppose that every voter knows, at zero cost, precisely how every government program will affect him or her.

In this imaginary world of costless information and ambitious legislators, each representative would support programs designed to appeal to a majority of voters. Organized groups would play no special role. Each legislator would already know how every voter feels about every issue, and every voter would already know how every program will affect him or her. A world of costless information would have no lobbyists, no pressure groups seeking a particular legislative agenda. No voter would be more important than any other.

Now let us drop the assumption that information is costless but retain the assumption that each legislator's goal is to be reelected. Legislators no longer know how people in the district feel about each issue. Furthermore, voters may not be sure how particular programs will affect them. People can obtain this information, but it is costly.

In this more realistic world of costly information, special-interest groups suddenly play an important role. A legislator who does not know how elderly voters in his or her district feel about a certain issue may find a conversation with a representative of the American Association of Retired Persons (AARP) to be a useful source of information. A chat with a lobbyist for the Teamster's Union may reveal something about the views of union members in the district. These groups also may be able to influence voter preferences through speeches and through public information and political action efforts.

A legislator in a world of costly information thus relies on special-interest groups for information and for support. To ensure his or her reelection, the legislator might try to fashion a program that appeals not to a majority of individuals but to a coalition of special-interest groups capable of delivering the support of a majority of voters. These groups are likely to demand something in exchange for their support of a particular candidate; they are likely to seek special programs to benefit their members. The role of special-interest groups is thus inevitable, given the cost of information and the desire of politicians to win elections. In the real world, it is not individual voters who count but well-organized groups that can deliver the support of voters to a candidate.

Public choice theorists argue that the inevitable importance of special-interest groups explains many choices the public sector makes. Consider, for example, the fact noted earlier in this chapter that a great many U.S. transfer payments go to groups, many of whose members are richer than the population as a whole. In the public choice perspective, the creation of a federal transfer program, even one that is intended to help poor people, will lead to competition among interest groups to be at the receiving end of the transfers. To win at this competition, a group needs money and organization—things poor people are not likely to have. In the competition for federal transfers, then, it is the nonpoor who often win.

The perception of growing power of special-interest groups in the United States has led to proposals for reform. One is the imposition of term limits, which restrict the number of terms a legislator can serve. Term limits were first established in Colorado in 1990; California and Oklahoma established term limits the same year. Subsequently, 18 other states adopted them. They have been found unconstitutional in four State Supreme Courts (Massachusetts, Oregon, Washington, and Wyoming). They have been repealed by the state legislatures of Idaho and Utah. Thus, term limits now apply in 15 states.[1]

One argument for term limits from the public choice perspective is that over time, incumbent legislators establish such close relationships with interest groups that they are virtually assured reelection; limiting terms may weaken these relationships and weaken special interests. The Supreme Court ruled in 1995 that individual states could not impose term limits on members of Congress. If such limits are to prevail at the federal level, a constitutional amendment will be required.

Arguments against the term limits approach include the fact that term limits automatically remove experienced legislators who could be very effective. They also restrict voter choice.

A second type of reform effort is a proposal that campaigns for seats in Congress be federally funded. If candidates did not need to seek funding from special interests, the influence of these groups would wane.

KEY TAKEAWAYS

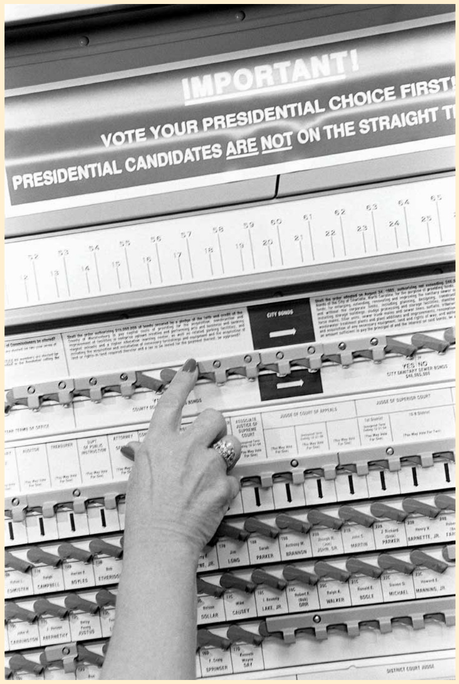
- Public interest theory examines government as an institution that seeks to maximize public well-being or net social benefit. It assumes government will seek the efficient solution to market failure problems.
- Public choice theory assumes that individuals engage in rent-seeking behavior by pursuing their self-interest in their dealings with the public sector; they continue to try to maximize utility or profit.
- It may be rational for eligible voters to abstain from voting, according to the public choice theory.
- Public choice theory suggests that politicians seeking reelection will try to appeal to coalitions of special-interest groups.

TRY IT!

Here is a list of possible explanations for government programs and policies. In each case, identify whether the explanation reflects the public interest theory or the public choice theory of government action.

1. "It is possible to explain much government activity by investigating the public's demand for government services, but one should not ignore the incentives for increased supply of government services."
2. "Through careful application of cost-benefit analysis, we can identify the amount of a public good that should be provided by the government."
3. "The determination of what are merit or demerit goods is inherently political rather than scientific and more often than not can be traced to the efforts of groups with an ax to grind or some private motive to pursue."
4. "While it is possible that policy makers follow some well-reasoned-out application of ability-to-pay or benefit-received principles, it is more credible to recognize that many of the taxes in this country reflect the fact that groups find it in their interest to organize to get tax burdens shifted to others."
5. "It is in the public interest to correct the market failure caused by monopoly firms. Therefore, it behooves us to do so."

Case in Point: The Presidential Election of 2000



© 2010 Jupiterimages Corporation

Public opinion polls on the eve of the election between George W. Bush and Al Gore showed the race to be a toss-up. Ordinarily, one might expect this to produce a large turnout. But barely more than half—50.7%—of registered voters went to the polls.

The 2000 election provides an illustration of the concept of rational abstention. It also illustrates another point made in the text. If an election is close, the outcome is likely to be determined in the courts.

Florida, with its 25 electoral votes, proved to be the decisive state. The winner of that state's electoral votes would win the presidency. The outcome in that state was not determined until late November, when Florida's Secretary of State, Republican Katherine Harris, declared George Bush the winner by a few hundred votes. Mr. Gore took the case to court. The Florida State Supreme Court ordered a recount.

The recounting process proved to be one of the most bizarre chapters in American political history. Thousands of lawyers descended on the state. Each ballot in key counties was scrutinized in an effort to determine which candidate each voter "intended" to choose. Chads, the small pieces of paper that are removed from a punch-card ballot, turned out to be of crucial importance. "Hanging chads," which occurred when the ballot was not thoroughly punched and which literally remained hanging from the ballot, prevented a ballot from being counted by the state's electric counting machines. The Florida's Supreme Court ruled that the roughly 170,000 ballots that had been discarded by the machines because they were not properly punched had to be re-examined.

As the recounting went on, other controversies arose. Pursuant to Florida law, Ms. Harris had ordered County Clerks to remove ex-felons from their registered voter lists. One clerk, seeing her own name on the list, refused to remove the names. Ms. Harris had come up with a list of 57,700 ex-felons for her "scrub list." The precise number of voters removed is not known. *Harper's Magazine* columnist Greg Palast charges that 90% of the voters on the scrub list were not, in fact, ex-felons. He notes that they were, however, black—and likely to vote Democratic—90% of ex-felons who are allowed to vote vote Democratic.

In the end, the case went to the United States Supreme Court. The Court decided, by a single vote, that Ms. Harris's certification of the outcome would stand, and George Bush became the president-elect of the United States.

All elections have stories of irregularities. The 2000 election was certainly no exception. What made it different was that the outcome came down to the votes in a single state. The official tally in Florida had Mr. Bush with 2,912,790 and Mr. Gore with 2,912,253. What was the "real" outcome? No one will ever know.

Sources: Florida Secretary of State, John Fund, "Vote Early and…" The Wall Street Journal, December 12, 2001; Greg Palast, "The Great Florida Ex-Con Game," Harper's Magazine, March 1, 2002; and U. S. Supreme Court, George W. Bush et al. vs. Albert Gore, Jr. et al., December 12, 2000.

ANSWER TO TRY IT! PROBLEM

Statements (2) and (5) reflect a public interest perspective. Statements (1), (3), and (4) reflect a public choice perspective.

4. REVIEW AND PRACTICE

Summary

In this chapter we examined the role of the public sector in the market economy. Since 1929, both the size and scope of government activities in the market have expanded considerably in the United States.

People demand government participation in three areas of economic activity. First, people may want correction of market failure involving public goods, external costs and benefits, and inefficient allocation created by imperfect competition. In each case of market failure, the shift from an inefficient allocation to an efficient one has the potential to eliminate or reduce deadweight losses. Second, people may seek government intervention to expand consumption of merit goods and to reduce consumption of demerit goods. Third, people often want government to participate in the transfer of income. Programs to transfer income have grown dramatically in the United States within the past few decades. The bulk of transfer payment spending is not means-tested.

Government activity is financed primarily by taxes. Two principles of taxation are the ability-to-pay principle, which holds that tax payments should rise with income, and the benefits-received principle, which holds that tax payments should be based on the benefits each taxpayer receives. Taxes may be regressive, proportional, or progressive. The major types of taxes in the United States are income taxes, sales and excise taxes, and property taxes. Economists seek to determine who bears the burden of a tax by examining its incidence. Taxes may be borne by buyers or sellers, depending on the relative elasticities of demand and supply.

Two broad perspectives are used to examine choices in the public sector. One is the public interest approach, which uses cost-benefit analysis to find the efficient solution to resource allocation problems. It assumes that the goal of the public sector is to maximize net social benefits. Cost-benefit analysis requires the estimation of benefits and costs that are not revealed in the marketplace. The second approach to the analysis of the public sector is public choice theory, which assumes utility-maximizing and rent-seeking behavior on the part of participants in the public sector and those trying to influence it. We examined two insights stemming from public choice theory: the problem of rational abstention from voting and the role of special interests.

CONCEPT PROBLEMS

1. Identify each of the following government programs as efforts to correct market failure, to promote or discourage the consumption of merit or demerit goods, or to transfer income.
    a. Head Start, a preschool program for low-income children
    b. Sports leagues for children sponsored by local governments
    c. A program to limit air pollution generated by power plants
    d. Species preservation efforts by the government
2. Public Broadcasting System (PBS) stations regularly solicit contributions from viewers. Yet only about 11% of these viewers, who on average have much higher incomes than the rest of the population, ever contribute. Why?
3. Do you expect to benefit from the research efforts sponsored by the American Cancer Society? Do you contribute? If you answered "Yes," then "No," does this make you a free rider?
4. Suppose the population of the United States increases. What will happen to the demand for national defense? What will happen to the efficient quantity of defense?
5. How could a program that redistributes income from rich to poor be considered a public good?
6. We noted that local governments typically supply tennis courts but not bowling alleys. Can you give a public choice explanation for this phenomenon? How about a public interest explanation?
7. Find out the turnout at the most recent election for student body president at your school. Does the turnout indicate student apathy?
8. Some welfare programs reduce benefits by $1 for every $1 that recipients earn; in effect, this is a tax of 100% on recipient earnings. Who pays the tax?
9. Suppose the quality of elementary education is a public good. How might we infer the demand for elementary school quality from residential property values?
10. V.I. Lenin, founder of the former Soviet Union, wrote that "the State is a machine for the oppression of one class by another." Explain whether Lenin's view typifies the public interest or the public choice school of public sector choice.
11. Sugar prices in the United States are several times higher than the world price of sugar. This disparity results from a federal government program that keeps enough foreign-produced sugar out of the United States to hold U.S. sugar prices at a high level. The program raises the price of all sweetened foods produced in the United States; it boosts food costs for the average household by more than a hundred dollars per year. Who benefits from the program? Why do you suppose it exists?
12. The table on federal income tax rates facing various income groups suggests that the marginal tax rate in the United States has fallen since the 1993–1996 period used in the study of marginal tax rates and labor supply discussed in the Case in Point essay. What would your prediction be as to how this reduction in the marginal tax rate would affect the quantity of labor supplied in the United States?
13. Given that we cannot have a perfectly accurate count of the votes in any election, is there any point in having elections at all?

NUMERICAL PROBLEMS

1. In an effort to beautify their neighborhood, four households are considering leasing a small section of vacant land for a park. For a monthly leasing fee, the owner of the vacant land is willing to arrange for some of the maintenance and to make the park available only to the four households. The demand curves for the four households (A, B, C, and D) wanting parkland are as follows (all demand curves are linear):

	Acres of Parkland Demanded per Month			
	A	B	C	D
$100	0	0	0	0
$75	1	0	0	0
$50	2	1⅓	0	0
$25	3	2⅔	2	0
$0	4	4	4	1

   Draw the demand curves for the four neighbors, and show the neighborhood demand curve for parkland.

2. Suppose the owner of the vacant land will provide for and maintain a neighborhood park at a fee of $125 per acre; the neighbors may lease up to 5 acres of land per month. Add this information to the graph you drew in Problem 1, and show the efficient solution. Are the neighbors likely to achieve this solution? Explain the problems involved in achieving it.
3. The perfectly competitive blank compact disc industry is in long-run equilibrium, selling blank discs for $5 apiece. Now the government imposes an excise tax of $2 per disc produced.
   a. Show what happens to the price and output of discs in the short run.
   b. Now show the impact in the long run.
   c. Who pays the tax? (Note: Show quantities as Q_1, Q_2, etc.)
4. A monopoly firm has just taken over the blank compact-disc industry. There have been technological advances that have lowered production cost, but the monopoly firm charges a price greater than average total cost, even in the long run. As it turns out, the firm is still selling compact discs for $5. The government imposes an excise tax of $2 per disc produced.
   a. What happens to price?
   b. What happens to output?
   c. Compare your results to your answer in Problem 3 and explain.
5. The following hypothetical data give annual spending on various goods and services for households at different income levels. Assume that an excise tax on any of these would, in the long run, be shifted fully to consumers.

Income range	Average income	Food	Clothing	Entertainment
$0–$25,000	$20,000	$5,000	$1,000	$500
$25,000–$50,000	$40,000	$8,000	$2,000	$2,000
$50,000–$75,000	$65,000	$9,750	$3,250	$5,200
$75,000–$100,000	$80,000	$10,000	$4,000	$8,000
> $100,000	$200,000	$16,000	$10,000	$30,000

   Determine whether a tax on any of the following goods would be progressive, proportional, or regressive.
   a. Food.
   b. Clothing.
   c. Entertainment.

ENDNOTES

1. "Legislative Term Limits: An Overview," National Conference of State Legislatures, April 22, 2005.

CHAPTER 16
Antitrust Policy and Business Regulation

START UP: THE PLASTIC WAR

The $2.5 trillion market for credit and debit cards received a major jolt in 2004 when the U.S. Supreme Court let stand a lower court ruling that Visa and MasterCard had violated the nation's antitrust laws by prohibiting banks who issued Visa and/or MasterCard from issuing Discover or American Express cards. The court found that, rather than competing with each other, Visa and MasterCard had cooperated with each other by increasing their "intercharge fees," the fees credit card companies charge to merchants who accept credit cards for payment, in lock-step. And, by locking Discover and American Express out of many markets, Visa and MasterCard were guilty of anti-competitive behavior.

The court's ruling spelled major trouble for Visa and MasterCard. Under U.S. law, a competitor that has been damaged by the anticompetitive practices of dominant firms can recover triple the damages that actually occurred. Rivals Discover and American Express filed suits against Visa and MasterCard. In 2008, American Express reached an agreement with MasterCard for a settlement of $1.8 billion. That followed a 2007 settlement with Visa for $2.1 billion. Together, the two agreements represented the largest judgments in America's antitrust history. Discover's $6 billion suit was still pending in mid-2008.[1] The government's case against Visa and MasterCard illustrates one major theme of this chapter.

In this chapter we will examine some of the limits government imposes on the actions of private firms. The first part of the chapter considers the effort by the U.S. government to limit firms' monopoly power and to encourage competition in the marketplace. The second part looks at those policies in the context of the global economy. We will also examine efforts to modify antitrust policy to make the U.S. economy more competitive internationally. In the third part of the chapter we will consider other types of business regulation, including those that seek to enhance worker and consumer safety, as well as deregulation efforts over the last 30 years.

1. ANTITRUST LAWS AND THEIR INTERPRETATION

LEARNING OBJECTIVES

1. **Define antitrust policies and tell when and why they were introduced in the United States.**
2. **Discuss highlights in the history of antitrust policies in the United States, focusing on major issues.**
3. **Explain the guidelines the Justice Department uses in dealing with mergers.**

antitrust policy

Policy by which government attempts to prevent the acquisition and exercise of monopoly power and to encourage competition in the marketplace.

In the decades after the Civil War, giant corporations and cartels began to dominate railroads, oil, banking, meat packing, and a dozen other industries. These businesses were led by entrepreneurs who, rightly or wrongly, have come to be thought of as "robber barons" out to crush their competitors, monopolize their markets, and gouge their customers. The term "robber baron" was associated with such names as J.P. Morgan and Andrew Carnegie in the steel industry, Philip Armour and Gustavas and Edwin Swift in meat packing, James P. Duke in tobacco, and John D. Rockefeller in the oil industry. They gained their market power through cartels and other business agreements aimed at restricting competition. Some formed *trusts*, a combination of corporations designed to consolidate, coordinate, and control the operations and policies of several companies. It was in response to the rise of these cartels and giant firms that antitrust policy was created in the United States. **Antitrust policy** refers to government attempts to prevent the acquisition and exercise of monopoly power and to encourage competition in the marketplace.

1.1 A Brief History of Antitrust Policy

The final third of the nineteenth century saw two major economic transitions. The first was industrialization—a period in which U.S. firms became far more capital intensive. The second was the emergence of huge firms able to dominate whole industries. In the oil industry, for example, Standard Oil of Ohio (after 1899, the Standard Oil Company of New Jersey) began acquiring smaller firms, eventually controlling 90% of U.S. oil-refining capacity. American Tobacco gained control of up to 90% of the market for most tobacco products, excluding cigars.

Public concern about the monopoly power of these giants led to a major shift in U.S. policy. What had been an economic environment in which the government rarely intervened in the affairs of private firms was gradually transformed into an environment in which government agencies took on a much more vigorous role. The first arena of intervention was antitrust policy, which authorized the federal government to challenge the monopoly power of firms head-on. The application of this policy, however, has followed a wandering and rocky road.

The Sherman Antitrust Act

The Sherman Antitrust Act of 1890 remains the cornerstone of U.S. antitrust policy. The Sherman Act outlawed contracts, combinations, and conspiracies in restraint of trade.

illegal per se

Actions taken by firms that are illegal in and of themselves without regard to the circumstances under which they occur.

rule of reason

Rule stating that whether or not a particular business practice is illegal depends on the circumstances surrounding the action.

An important issue in the interpretation of the Sherman Act concerns which actions by firms are **illegal per se**, meaning illegal in and of itself without regard to the circumstances under which it occurs. Shoplifting, for example, is illegal per se; courts do not inquire whether shoplifters have a good reason for stealing something in determining whether their acts are illegal. One key question of interpretation is whether it is illegal per se to control a large share of a market. Another is whether a merger that is likely to produce substantial monopoly power is illegal per se.

Two landmark Supreme Court cases in 1911 in which the Sherman Act was effectively used to break up Standard Oil and American Tobacco enunciated the **rule of reason**, which holds that whether or not a particular business practice is illegal depends on the circumstances surrounding the action. In both cases, the companies held dominant market positions, but the Court made it clear that it was their specific "unreasonable" behaviors that the breakups were intended to punish. In determining what was illegal and what was not, emphasis was placed on the conduct, not the structure or size, of the firms.

In the next 10 years, the Court threw out antitrust suits brought by government prosecutors against Eastman Kodak, International Harvester, United Shoe Machinery, and United States Steel. The Court determined that none of them had used unreasonable means to achieve their dominant positions in the industry. Rather, they had successfully exploited economies of scale to reduce costs below competitors' costs and had used reasonable means of competition to reap the rewards of efficiency.

The rule of reason suggests that "bigness" is no offense if it has been achieved through legitimate business practices. This precedent, however, was challenged in 1945 when the U.S. Court of Appeals ruled against the Aluminum Company of America (Alcoa). The court acknowledged that Alcoa had been able to capture over 90% of the aluminum industry through reasonable business practices. Nevertheless, the court held that by sheer size alone, Alcoa was in violation of the prohibition against monopoly.

In a landmark 1962 court case involving a proposed merger between United Shoe Machinery and the Brown Shoe Company, one of United's competitors, the Supreme Court blocked the merger because the resulting firm would have been so efficient that it could have undersold all of its competitors. The Court recognized that lower shoe prices would have benefited consumers, but chose to protect competitors instead.

The Alcoa case and the Brown Shoe case, along with many other antitrust cases in the 1950s and 1960s, added confusion and uncertainty to the antitrust environment by appearing to reinvoke the doctrine of per se illegality. In the government's case against Visa and MasterCard, the government argued successfully that the behavior of the two firms was a per se violation of the Sherman Act.

The Sherman Act also aimed, in part, to prevent **price-fixing**, in which two or more firms agree to set prices or to coordinate their pricing policies. For example, in the 1950s General Electric, Westinghouse, and several other manufacturers colluded to fix prices. They agreed to assign market segments in which one firm would sell at a lower price than the others. In 1961, the General Electric–Westinghouse agreement was declared illegal. The companies paid a multimillion-dollar fine, and their officers served brief jail sentences. In 2008, three manufactures of liquid crystal display panels—the flat screens used in televisions, cell phones, personal computers, and such—agreed to pay $585 million in fines for price fixing, with LG Display paying $400 million, Sharp Corporation paying $120 million, and Chunghwa Picture Tubes paying $65 million. The $400 million fine to LG is still less than the record single fine of $500 million paid in 1999 by F. Hoffman-LaRoche, the Swiss pharmaceutical company, in a case involving fixing prices of vitamin supplements.

price-fixing

Situation in which two or more firms agree to set prices or to coordinate their pricing policies.

Other Antitrust Legislation

Concerned about the continued growth of monopoly power, in 1914 Congress created the Federal Trade Commission (FTC), a five-member commission that, along with the antitrust division of the Justice Department, has the power to investigate firms that use illegal business practices.

In addition to establishing the FTC, Congress enacted new antitrust laws intended to strengthen the Sherman Act. The Clayton Act (1914) clarifies the illegal per se provision of the Sherman Act by prohibiting the purchase of a rival firm if the purchase would substantially decrease competition, and outlawing interlocking directorates, in which there are the same people sitting on the boards of directors of competing firms. More significantly, the act prohibits price discrimination that is designed to lessen competition or that tends to create a monopoly and exempts labor unions from antitrust laws.

The Sherman and Clayton acts, like other early antitrust legislation, were aimed at preventing mergers that reduce the number of firms in a single industry. The consolidation of two or more producers of the same good or service is called a **horizontal merger**. Such mergers increase concentration and, therefore, the likelihood of collusion among the remaining firms.

The Celler–Kefauver Act of 1950 extended the antitrust provisions of earlier legislation by blocking **vertical mergers**, which are mergers between firms at different stages in the production and distribution of a product if a reduction in competition will result. For example, the acquisition by Ford Motor Company of a firm that supplies it with steel would be a vertical merger.

horizontal merger

The consolidation of two or more producers of the same good or service.

vertical merger

Mergers between firms at different stages in the production and distribution of a product.

1.2 U.S. Antitrust Policy Today

The "bigness is badness" doctrine dominated antitrust policy from 1945 to the 1970s. But the doctrine always had its critics. If a firm is more efficient than its competitors, why should it be punished? Critics of the antitrust laws point to the fact that of the 500 largest companies in the United States in 1950, over 100 no longer exist. New firms, including such giants as Walmart, Microsoft, and Federal Express, have taken their place. The critics argue that the emergence of these new firms is evidence of the dynamism and competitive nature of the modern corporate scene.

There is no evidence to suggest, for example, that the degree of concentration across all industries has increased over the past 25 years. Global competition and the use of the internet as a marketing tool have increased the competitiveness of a wide range of industries. Moreover, critics of antitrust policy argue that it is not necessary that an industry be perfectly competitive to achieve the benefits of competition. It need merely be **contestable**—open to entry by potential rivals. A large firm may be able to prevent small firms from competing, but other equally large firms may enter the industry in pursuit of the high profits earned by the initial large firm. For example, Time Warner, primarily a competitor in

contestable

Open to entry by potential rivals.

the publishing and entertainment industries, has in recent years become a main competitor in the cable television market.

Currently, the Justice Department follows guidelines based on the Herfindahl–Hirschman Index (HHI). The HHI, introduced in an earlier chapter, is calculated by summing the squared percentage market shares of all firms in an industry, where the percentages are expressed as whole numbers (for example 30% would be expressed as 30). The higher the value of the index, the greater the degree of concentration. Possible values of the index range from 0 in the case of perfect competition to 10,000 ($= 100^2$) in the case of a monopoly.

Current guidelines stipulate that any industry with an HHI under 1,000 is unconcentrated. Except in unusual circumstances, mergers of firms with a postmerger index under 1,000 will not be challenged. The Justice Department has said it would challenge proposed mergers with a postmerger HHI between 1,000 and 1,800 if the index increased by more than 100 points. Industries with an index greater than 1,800 are deemed highly concentrated, and the Justice Department has said it would seek to block mergers in these industries if the postmerger index would increase by 50 points or more. Table 16.1 summarizes the use of the HHI by the Justice Department.

TABLE 16.1 The Herfindahl-Hirschman Index and Antitrust Policy

The Department of Justice (DOJ) and the Federal Trade Commission (FTC) have adopted the following guidelines for merger policy based on the Herfindahl-Hirschman Index.

If the postmerger Herfindahl-Hirschman Index is found to be…	then the Justice Department will likely take the following action.
Unconcentrated (<1,000)	No challenge
Moderately concentrated (1,000–1,800)	Challenge if postmerger index changes by more than 100 points.
Highly concentrated (>1,800)	Challenge if postmerger index changes by more than 50 points.

U.S. Department of Justice and Federal Trade Commission, 1992 Horizontal Merger Guidelines, issued April 2, 1992, revised April 8, 1997.

One difficulty with the use of the HHI is that its value depends on the definition of the market. With a sufficiently narrow definition of the market, even a highly competitive market could have an HHI close to the value for a monopoly. The late George Stigler commented on the difficulty in a fanciful discussion of the definition of the relevant market for cameras:

> *"Consider the problem of defining a market within which the existence of competition or some form of monopoly is to be determined. The typical antitrust case is an almost impudent exercise in economic gerrymandering. The plaintiff sets the market, at a maximum, as one state in area and including only aperture-priority SLR cameras selling between $200 and $250. This might be called J-Shermanizing the market, after Senator John Sherman. The defendant will in turn insist that the market be world-wide, and include not only all cameras, but all portrait artists and all transportation media, since a visit is a substitute for a picture. This might also be called T-Shermanizing the market, after the Senator's brother, General William Tecumseh Sherman. Depending on who convinces the judge, the concentration ratio will be awesome or trivial, with a large influence on the verdict."*[2]

Of course, the definition of the relevant market is not a matter of arbitrarily defining the market as absurdly narrow or broad. There are economic tests to determine the range of goods or services that should be included in a particular market. Consider, for example, the market for refrigerators. Given the relatively low cost of shipping refrigerators, the relevant area might encompass all of North America, given the existence of the North American Free Trade Agreement (NAFTA), which establishes a tariff-free trade zone including Canada, the United States, and Mexico. What sorts of goods should be included? Presumably, any device that is powered by electricity or by natural gas and that keeps things cold would qualify. Certainly, a cool chest that requires ice that people take on picnics would not be included. The usual test is the cross price elasticity of demand. If it is high between any two goods, then those goods are candidates for inclusion in the market.

Should the entire world be the geographic region for the market for refrigerators? That is an empirical question. If the cross price elasticities for refrigerator brands worldwide are high, then one would conclude that the world is the relevant geographical definition of the market.

In the 1980s both the courts and the Justice Department held that bigness did not necessarily translate into badness, and corporate mergers proliferated. In the period 1982–1989 there were almost

200 mergers and acquisitions of firms whose value exceeded $1 billion. The total value of these companies was nearly half a trillion dollars.

Megamergers continued in the 1990s and into the 21st-century. In 2000, there were 212 mergers valued at $1 billion or more and in 2006 nearly that many. Since then, merger activity has decreased, in part due to turmoil in financial markets.[3]

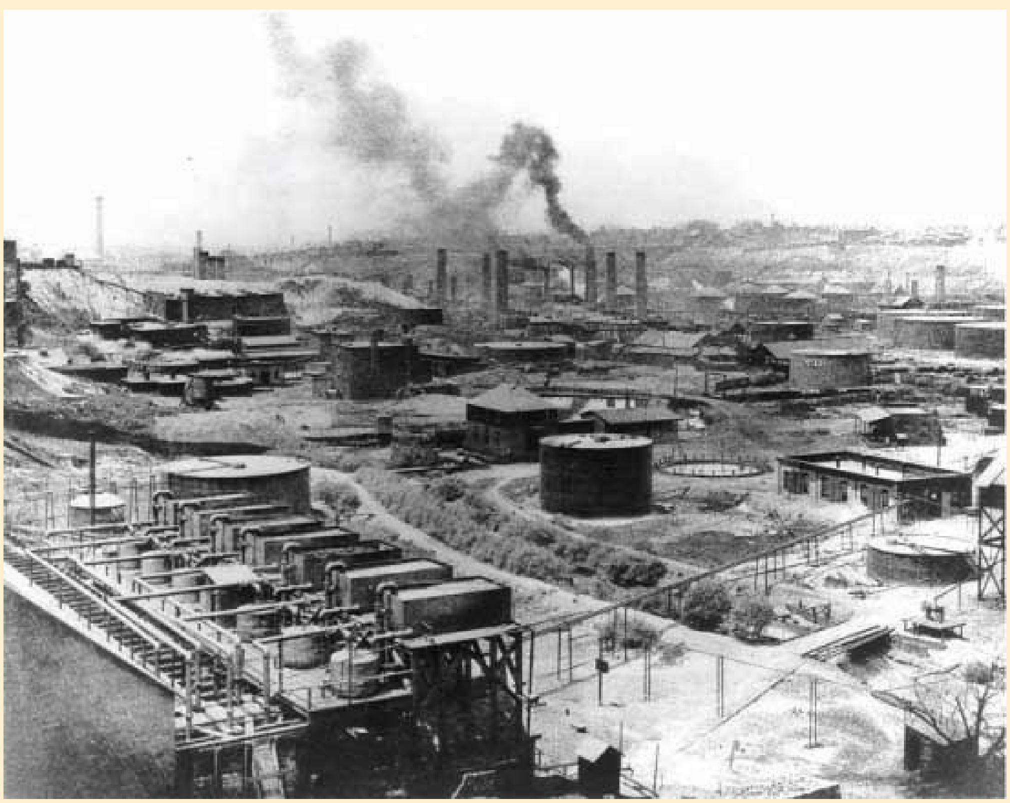
KEY TAKEAWAYS

- The government uses antitrust policies to maintain competitive markets in the economy.
- The Sherman Antitrust Act of 1890 and subsequent legislation defined illegal business practices, but these acts are subject to widely varying interpretations by government agencies and by the courts.
- Although price-fixing is illegal per se, most business practices that may lessen competition are interpreted under the rule of reason.
- The Justice Department and Federal Trade Commission use the Herfindahl-Hirschman Index to determine whether mergers should be challenged in particular industries.

TRY IT!

According to what basic principle did the U.S. Supreme Court find Eastman Kodak not guilty of violating antitrust laws? According to what basic principle did the Court block the merger of Brown Shoe Company and one of its competitors, United Shoe Machinery? Do you agree or disagree with the Court's choices?

Case in Point: Does Antitrust Policy Help Consumers?



The Department of Justice and the Federal Trade Commission spend a great deal of money enforcing U.S. antitrust laws. Firms defending themselves may spend even more.

The government's first successful use of the Sherman Act came in its action against Standard Oil in 1911. The final decree broke Standard into 38 independent companies. Did the breakup make consumers better off?

In 1899, Standard controlled 88% of the market for refined oil products. But, by 1911, its share of the market had fallen to 64%. New discoveries of oil had sent gasoline prices down in the years before the ruling. After the ruling, gasoline prices began rising. It does not appear that the government's first major victory in an antitrust case had a positive impact on consumers.

In general, antitrust cases charging monopolization take so long to be resolved that, by the time a decree is issued, market conditions are likely to have changed in a way that makes the entire effort seem somewhat frivolous. For example, the government charged IBM with monopolization in 1966. That case was finally dropped in 1982 when the market had changed so much that the original premise of the case was no longer valid. In 1998 the Department of Justice began a case against against Microsoft, accusing it of monopolizing the market for Internet browsers by bundling the browser with its operating system, Windows. A trial in 2000 ended with a judgment that Microsoft be split in two with one company having the operating system and another having applications. An appeals court overturned that decision a year later.

Actions against large firms such as Microsoft are politically popular. However, neither policy makers nor economists have been able to establish that they serve consumer interests.

We have seen that the Department of Justice and the Federal Trade Commission have a policy of preventing mergers in industries that are highly concentrated. But, mergers often benefit consumers by achieving reductions in cost. Perhaps the most surprising court ruling involving such a merger came in 1962 when the Supreme Court ruled that a merger in shoe manufacturing would achieve lower costs to consumers. The Court prevented the merger on grounds the new company would be able to charge a lower price than its rivals! Clearly, the Court chose to protect firms rather than to enhance consumer welfare.

What about actions against price-fixing? The Department of Justice investigates roughly 100 price-fixing cases each year. In many cases, these investigations result in indictments. Those cases would, if justified, result in lower prices for consumers. But, economist Michael F. Sproul, in an examination of 25 price-fixing cases for which adequate data were available, found that prices actually *rose* in the four years following most indictments.

Economists Robert W. Crandall and Clifford Winston have asked a very important question: Has all of this effort enhanced consumer welfare? They conclude that the Department of Justice and the Federal Trade Commission would best serve the economy by following a policy of benign neglect in cases of monopolization, proposed mergers, and efforts by firms to exploit technological gains by lowering price. The economists conclude that antitrust actions should be limited to the most blatant cases of price-fixing or mergers that would result in monopolies. In contrast, law professor Jonathan Baker argued in the same journal that such a minimalist approach could be harmful to consumer welfare. One argument he makes is that antitrust laws and their enforcement create a deterrence effect.

A recent paper by Orley Ashenfelter and Daniel Hosken analyzed the impact of five mergers in the consumer products industry that seemed to be most problematic for antitrust enforcement agencies. In four of the five cases prices rose following the mergers and in the fifth case the merger had little effect on price. While they do not conclude that this small study should be used to determine the appropriate level of government enforcement of antitrust policy, they state that those who advocate less intervention should note that the price effects were not negative, as they would have been if these mergers were producing cost decreases and passing them on to consumers. Those advocating more intervention should note that the price increases they observed after these mergers were not very large.

Sources: Orley Ashenfelter and Daniel Hosken, "The Effect of Mergers on Consumer Prices: Evidence from Five Selected Cases," National Bureau of Economic Research Working Paper 13859, March 2008; James B. Baker, "The Case for Antitrust Enforcement," Journal of Economic Perspectives, 17:4 (Fall 2003): 27–50; Robert W. Crandall and Clifford Winston, "Does Antitrust Policy Improve Consumer Welfare? Assessing the Evidence," Journal of Economic Perspectives, 17:4 (Fall 2003): 3–26; Michael F. Sproul, "Antitrust and Prices," Journal of Political Economy, 101 (August 1993): 741–54.

ANSWER TO TRY IT! PROBLEM

In the case of Eastman Kodak, the Supreme Court argued that the rule of reason be applied. Even though the company held a dominant position in the film industry, its conduct was deemed reasonable. In the proposed merger between United Shoe Machinery and Brown Shoe, the court clearly chose to protect the welfare of firms in the industry rather than the welfare of consumers.

2. ANTITRUST AND COMPETITIVENESS IN A GLOBAL ECONOMY

LEARNING OBJECTIVES

1. **Define joint ventures and explain the evolution of U.S. antitrust policy towards them.**
2. **Discuss other antitrust policy changes that relate to U.S. firms competing with foreign firms.**

In the early 1980s, U.S. imports from foreign firms rose faster than U.S. exports. In 1986 the trade deficit reached a record level at that time. Antitrust laws played a relatively minor role in increasing the deficit, but business interests and politicians pressed for the relaxation of antitrust policy in order to make U.S. firms more competitive against multinational companies headquartered in other countries.

Antitrust enforcement was altered in the late 1980s so that horizontally competitive U.S. firms could cooperate in research and development (R&D) ventures aimed at innovation, cost-cutting technological advances, and new product development. In an antitrust context, **joint ventures** refer to cooperative arrangements between two or more firms that otherwise would violate antitrust laws. Proponents of the change argued that foreign multinational firms were not subject to stringent antitrust restrictions and therefore had a competitive advantage over U.S. firms. The International Competition Policy Advisory Committee (ICPAC) was formed in the Department of Justice in 1997 in recognition of the dramatic increases in both international commerce and international anticompetitive activity. Composed of a panel of business, industrial relations, academic, economic, and legal experts, ICPAC is to provide advice and information to the department on international antitrust issues such as transnational cartels and international anticompetitive business practices.

joint ventures

Cooperative arrangements between two or more firms that otherwise would violate antitrust laws.

2.1 Cooperative Ventures Abroad

Policymakers who revised U.S. antitrust restrictions on joint ventures pointed out that Japanese and European firms are encouraged to cooperate and to collude not only in basic R&D projects, but in production and marketing as well.

The evidence is difficult to interpret, but in Japan, for example, a substantial percentage of research projects are sponsored jointly by firms in the same market. Moreover, the evidence is fairly clear that Japan allows horizontal consolidations and mergers in moderately concentrated markets where antitrust policy would be applied in the United States. Mergers that create substantial monopoly power in Japan are not typically prosecuted by the government.

In Europe, the potential competitive threat to U.S. firms is twofold. First, as the European Union (EU) moved toward economic unification in 1992, it relaxed antitrust enforcement for mergers between firms in different nations, even though they would become a single transnational firm in the near future. In 1984, for example, the European Community (EC), the forerunner of the EU, adopted a regulation that provided blanket exemptions from antitrust provisions against collusion in R&D for firms whose total market share did not exceed 20%. This exemption included horizontal R&D and extended to production and distribution to the point of final sale. Moreover, firms that had a market share greater than 20% could apply for an exemption based on a case-by-case examination.

The U.S. government has relaxed antitrust restrictions in some cases in an effort to make domestic firms more competitive in global competition. For example, producers of semiconductors were allowed to form a research consortium, Sematech, in order to promote the U.S. semiconductor industry. This type of joint venture was formerly prohibited. Sematech has since created the International Sematech Manufacturing Initiative (ISMI), a wholly owned subsidiary dedicated to improve the productivity and cost performance of equipment and manufacturing operations well beyond a narrowly defined semiconductor industry. Its membership includes both domestic and foreign firms, and they collectively represent half of the world's integrated circuit (semiconductor and microchip) production. In this case, we see the U.S. government is supporting cooperation among multinational and international firms ostensibly for product improvement. One suspects, however, that member firms gain a competitive advantage over non-member firms wherever in the world they are located.

2.2 Antitrust Policy and U.S. Competitiveness

In the 1980s Congress passed several laws that relaxed the antitrust prohibition against cooperation among U.S. firms, including the National Cooperative Research Act of 1984 (NCRA) and the Omnibus Trade and Competitiveness Act (OTCA).

The NCRA provided a simple registration procedure for joint ventures in R&D. The NCRA protects members of registered joint ventures from punitive antitrust penalties if the venture is later found to illegally reduce competition or otherwise act in restraint of trade. Between 1984 and 1990 over 200 research joint ventures were registered, substantially more than were formed over the same period within the EC.

Congress passed the OTCA in 1988. The OTCA made unfair methods of competition by foreign firms and importers punishable under the U.S. antitrust laws. It also changed the wording of existing laws concerning "dumping" (selling below cost) by foreign firms. In the past, a domestic firm that claimed injury from a foreign competitor had to prove that the foreign firm was "undercutting" the U.S. market prices. The OTCA changed this provision to the much less restrictive term "underselling" and specifically stated that the domestic firm did not have to prove predatory intent. This legislation opened the door for U.S. competitors to use antitrust laws to *prevent* competition from foreigners, quite the opposite of the laws' original purpose. Dumping is discussed further in a later chapter.

In another policy shift, the Justice Department announced in 1988 that the rule of reason would replace per se illegality in analysis of joint ventures that would increase U.S. competitiveness. The Justice Department uses the domestic guidelines and the Herfindahl–Hirschman Index to determine whether a proposed joint venture would increase concentration and thereby lessen competition. In making that assessment, the Justice Department also looks at (1) whether the firms directly compete in other markets, (2) the possible impact on vertical markets, and (3) whether any offsetting efficiency benefits outweigh the anticompetitiveness of the joint venture. Although mergers between two firms in a moderately or highly concentrated industry are prohibited, joint ventures between them may be allowed.

The major antitrust issues to be resolved in the first decade of the twenty-first century go beyond joint R&D ventures. The World Trade Organization, the international organization created in 1995 to supervise world trade, has established a group to study issues relating to the interaction between trade and competition policy, including anticompetitive practices. Nations currently have quite different antitrust laws, as the Case in Point in this section illustrates. The United States has argued against any internationalization of antitrust issues that would reduce its ability to apply U.S. laws. On the other hand, the United States, via the 1994 International Antitrust Enforcement Assistance Act, is negotiating bilateral agreements that allow antitrust agencies in different countries to exchange information for use in antitrust enforcement. The issue of how to deal with anticompetitive practices on a worldwide basis remains unresolved, and this area of antitrust practice and policy will be closely watched and studied by economists.

KEY TAKEAWAYS

- Increased imports in the last 25 years have led to a rethinking of American antitrust policy.
- One response by the U.S. to international competition is the encouragement of joint ventures.
- U.S. firms that have been "undersold" by foreign firms can charge those firms with "dumping."
- The World Trade Organization is studying the interactions of trade, competition, and antitrust issues.

TRY IT!

Suppose that long-distance companies in the United States form a joint venture to explore alternative technologies in telephone services. Would such an effort suggest any danger of collusion? Would it be likely to be permitted?

Case in Point: The United States and the European Union—Worlds Apart

© 2010 Jupiterimages Corporation

The European Union's initial reaction to the proposed merger of Boeing and McDonnell Douglas in 1997 was to threaten to impose tariffs on Boeing planes entering the continent if the deal went through. The issue brought the United States and its European partners to the brink of a trade war.

Then President Bill Clinton responded to the EU's threat saying, "I'm concerned about what appears to be the reasons for the objection to the Boeing-McDonnell Douglas merger by the European Union, and we have some options ourselves when actions are taken in this regard." The president seemed to be suggesting retaliatory trade sanctions, such as U.S. tariffs on European-made planes.

At the last minute, the EU allowed the merger on two conditions: that Boeing give up its exclusive supply deals and agreed to license to its competitors (meaning Airbus) McDonnell technology that had been developed with U.S. government support.

In the press, the incident was reported as an incipient trade war. Europe was trying to protect its own airline industry; the United States its own. According to New York University economist Eleanor Fox, though, the dispute stemmed not from countries trying to protect their own companies but from differing antitrust laws.

Ms. Fox argues that U.S. antitrust law is consumer oriented. The question for the Federal Trade Commission was whether the merger made consumers worse off by raising the price of jets to airlines. The FTC reasoned that McDonnell Douglas had no reasonable chance of making and selling new fleets on its own and thus did not constitute a competitive force in the marketplace. With McDonnell Douglas deemed competitively insignificant, the merger was permissible.

However, European Union antitrust laws consider not only consumers but also unfair competitive advantages of dominant firms. Because Boeing held 20-year exclusive contracts with three airlines that represent more than 10% of the market for airline manufacture, the merger magnified Boeing's competitive advantage over other firms (primarily Airbus) that sell aircraft. The conditions that the EU impose thus made the merger subject to its antitrust laws.

The policy difference is fundamental. Americans argue that they seek to protect competition, while the EU protects competitors—even if consumers suffer as a result. The *Economist*, a British newsmagazine, reports American antitrust policy makers tend to rely on market forces to dampen monopoly power and argue that relying on regulation may tend to diminish innovation and, in the long run, competition. Europeans argue that regulation is necessary in order to ensure that all firms have a reasonable chance to compete.

The difference in the two approaches to antitrust is vividly illustrated in the treatment of Microsoft by the United States and by the European Union. While the United States initially attempted to prosecute Microsoft for violating the Sherman Act by bundling Internet Explorer with its Windows software, it has since permitted it. The European Union has come down very hard on Microsoft, fining it €1.4 billion ($2.2 billion) and ordering the firm to supply firms using Windows the complete documentation of the system. U.S. authorities argue that such restrictions make Microsoft a less innovative company and argue that the computer market is a highly competitive one as it is and that the imposition of a regulatory burden risks stifling the competition that exists.

Sources: Eleanor M. Fox, "Antitrust Regulation Across International Borders," The Brookings Review, 16(1) (Winter 1998): 30–32; "Oceans Apart," Economist, May 1, 2008, 387(8578): 78–79.

ANSWER TO TRY IT! PROBLEM

A joint venture between competing long-distance companies carries the danger that they may end up colluding. It is also possible that only some long-distance firms would be involved to the exclusion of other rival firms, as happened in the joint venture between General Motors and Toyota. On the other hand, the venture might be allowed in the U.S. under the notion that the firms might need to cooperate to face global competition. Another consideration is that technological change in this industry is occurring so rapidly that competitors can emerge from anywhere. Cable companies, internet providers, and cellular-phone companies all compete with regular telephone companies.

3. REGULATION: PROTECTING PEOPLE FROM THE MARKET

LEARNING OBJECTIVES

1. **Compare the public interest and public choice theories of regulation.**
2. **Discuss the costs and benefits of consumer protection laws.**
3. **Discuss the pros and cons of the trend toward deregulation over the last quarter century.**

regulation

Government power to control the choices of private firms or individuals.

Antitrust policies are primarily concerned with limiting the accumulation and use of market power. Government **regulation** is used to control the choices of private firms or individuals. Regulation may constrain the freedom of firms to enter or exit markets, to establish prices, to determine product design and safety, and to make other business decisions. It may also limit the choices made by individuals.

In general terms, there are two types of regulatory agencies. One group attempts to protect consumers by limiting the possible abuse of market power by firms. The other attempts to influence business decisions that affect consumer and worker safety. Regulation is carried out by more than 50 federal agencies that interpret the applicable laws and apply them in the specific situations they find in real-world markets. Table 16.2 lists some of the major federal regulatory agencies, many of which are duplicated at the state level.

TABLE 16.2 Selected Federal Regulatory Agencies and Their Missions

Financial Markets	
Federal Reserve Board	Regulates banks and other financial institutions
Federal Deposit Insurance Corporation	Regulates and insures banks and other financial institutions
Securities and Exchange Commission	Regulates and requires full disclosure in the securities (stock) markets
Commodity Futures Trading Commission	Regulates trading in futures markets
Product Markets	
Department of Justice, Antitrust Division	Enforces antitrust laws
Federal Communications Commission	Regulates broadcasting and telephone industries
Federal Trade Commission	Focuses efforts on consumer protection, false advertising, and unfair trade practices
Federal Maritime Commission	Regulates international shipping
Surface Transportation Board	Regulates railroads, trucking, and noncontiguous domestic water transportation
Federal Energy Regulatory Commission	Regulates pipelines
Health and Safety	
Occupational Health and Safety Administration	Regulates health and safety in the workplace
National Highway Traffic Safety Administration	Regulates and sets standards for motor vehicles
Federal Aviation Administration	Regulates air and traffic aviation safety
Food and Drug Administration	Regulates food and drug producers; emphasis on purity, labeling, and product safety
Consumer Product Safety Commission	Regulates product design and labeling to reduce risk of consumer injury
Energy and the Environment	
Environmental Protection Agency	Sets standards for air, water, toxic waste, and noise pollution
Department of Energy	Sets national energy policy
Nuclear Regulatory Commission	Regulates nuclear power plants
Corps of Engineers	Sets policies on construction near rivers, harbors, and waterways
Labor Markets	
Equal Employment Opportunity Commission	Enforces antidiscrimination laws in the workplace
National Labor Relations Board	Enforces rules and regulations governing contract bargaining and labor relations between companies and unions

3.1 Theories of Regulation

Competing explanations for why there is so much regulation range from theories that suggest regulation protects the public interest to those that argue regulation protects the producers or serves the interests of the regulators. The distinction corresponds to our discussion in the last chapter of the public interest versus the public choice understanding of government policy in general.

The Public Interest Theory of Regulation

The public interest theory of regulation holds that regulators seek to find market solutions that are economically efficient. It argues that the market power of firms in imperfectly competitive markets must be controlled. In the case of natural monopolies (discussed in an earlier chapter), regulation is viewed as necessary to lower prices and increase output. In the case of oligopolistic industries, regulation is often advocated to prevent cutthroat competition.

The public interest theory of regulation also holds that firms may have to be regulated in order to guarantee the availability of certain goods and services—such as electricity, medical facilities, and telephone service—that otherwise would not be profitable enough to induce unregulated firms to provide them in a given community. Firms providing such goods and services are often granted licenses and franchises that prevent competition. The regulatory authority allows the firm to set prices above average cost in the protected market in order to cover losses in the target community. In this way, the firms are allowed to earn, indeed are guaranteed, a reasonable rate of return overall.

Proponents of the public interest theory also justify regulation of firms by pointing to externalities, such as pollution, that are not taken into consideration when unregulated firms make their decisions. As we have seen, in the absence of property rights that force the firms to consider all of the costs and benefits of their decisions, the market may fail to allocate resources efficiently.

The Public Choice Theory of Regulation

capture theory of regulation

Theory stating that government regulations often end up serving the regulated firms rather than their customers.

The public interest theory of regulation assumes that regulations serve the interests of consumers by restricting the harmful actions of business firms. That assumption, however, is now widely challenged by advocates of the public choice theory of regulation, which rests on the premise that all individuals, including public servants, are driven by self-interest. They prefer the **capture theory of regulation**, which holds that government regulations often end up serving the regulated firms rather than their customers.

Competing firms always have an incentive to collude or operate as a cartel. The public is protected from such collusion by a pervasive incentive for firms to cheat. Capture theory asserts that firms seek licensing and other regulatory provisions to prevent other firms from entering the market. Firms seek price regulation to prevent price competition. In this view, the regulators take over the role of policing cartel pricing schemes; individual firms in a cartel would be incapable of doing so themselves.

Because it is practically impossible for the regulatory authorities to have as much information as the firms they are regulating, and because these authorities often rely on information provided by those firms, the firms find ways to get the regulators to enforce regulations that protect profits. The regulators get "captured" by the very firms they are supposed to be regulating.

In addition to its use of the capture theory, the public choice theory of regulation argues that employees of regulatory agencies are not an exception to the rule that people are driven by self-interest. They maximize their own satisfaction, not the public interest. This insight suggests that regulatory agencies seek to expand their bureaucratic structure in order to serve the interests of the bureaucrats. As the people in control of providing government protection from the rigors of the market, bureaucrats respond favorably to lobbyists and special interests.

Public choice theory views the regulatory process as one in which various groups jockey to pursue their respective interests. Firms might exploit regulation to limit competition. Consumers might seek lower prices or changes in products. Regulators themselves might pursue their own interests in expanding their prestige or incomes. The abstract goal of economic efficiency is unlikely to serve the interest of any one group; public choice theory does not predict that efficiency will be a goal of the regulatory process. Regulation might improve on inefficient outcomes, but it might not.

3.2 Consumer Protection

Every day we come into contact with regulations designed to protect consumers from unsafe products, unscrupulous sellers, or our own carelessness. Seat belts are mandated in cars and airplanes; drivers must provide proof of liability insurance; deceptive advertising is illegal; firms cannot run "going out of business" sales forever; electrical and plumbing systems in new construction must be inspected and approved; packaged and prepared foods must carry certain information on their labels; cigarette packages must warn users of the dangers involved in smoking; gasoline stations must prevent gas spillage; used-car odometers must be certified as accurate. The list of regulations is seemingly endless.

There are often very good reasons behind consumer protection regulation, and many economists accept such regulation as a legitimate role and function of government agencies. But there are costs as well as benefits to consumer protection.

The Benefits of Consumer Protection

Consumer protection laws are generally based on one of two conceptual arguments. The first holds that consumers do not always know what is best for them. This is the view underlying government efforts to encourage the use of merit goods and discourage the use of demerit goods. The second suggests that consumers simply do not have sufficient information to make appropriate choices.

Laws prohibiting the use of certain products are generally based on the presumption that not all consumers make appropriate choices. Drugs such as cocaine and heroin are illegal for this reason.

Children are barred from using products such as cigarettes and alcohol on grounds they are incapable of making choices in their own best interest.

Other regulations presume that consumers are rational but may not have adequate information to make choices. Rather than expect consumers to determine whether a particular prescription drug is safe and effective, for example, federal regulations require the Food and Drug Administration to make that determination for them.

The benefit of consumer protection occurs when consumers are prevented from making choices they would regret if they had more information. A consumer who purchases a drug that proves ineffective or possibly even dangerous will presumably stop using it. By preventing the purchase in the first place, the government may save the consumer the cost of learning that lesson.

One problem in assessing the benefits of consumer protection is that the laws themselves may induce behavioral changes that work for or against the intent of the legislation. For example, requirements for childproof medicine containers appear to have made people more careless with medicines. Requirements that mattresses be flame-resistant may make people more careless about smoking in bed. In some cases, then, the behavioral changes attributed to consumer protection laws may actually worsen the problem the laws seek to correct.

An early study on the impact of seat belts on driving behavior indicated that drivers drove more recklessly when using seat belts, presumably because the seat belts made them feel more secure.[4] A recent study, however, found that this was not the case and suggests that use of seat belts may make drivers more safety-conscious.[5]

In any event, these "unintended" behavioral changes can certainly affect the results achieved by these laws.

The Cost of Consumer Protection

Regulation aimed at protecting consumers can benefit them, but it can also impose costs. It adds to the cost of producing goods and services and thus boosts prices. It also restricts the freedom of choice of individuals, some of whom are willing to take more risks than others.

Those who demand, and are willing to pay the price for, high-quality, safe, warranted products can do so. But some argue that people who demand and prefer to pay (presumably) lower prices for lower-quality products that may have risks associated with their use should also be allowed to exercise this preference. By increasing the costs of goods, consumer protection laws may adversely affect the poor, who are forced to purchase higher-quality products; the rich would presumably buy higher-quality products in the first place.

To assess whether a particular piece of consumer protection is desirable requires a careful look at how it stacks up against the marginal decision rule. The approach of economists is to attempt to determine how the costs of a particular regulation compare to its benefits.

Economists W. Mark Crain and Thomas D. Hopkins estimated the cost of consumer protection regulation in 2001 and found that the total cost was $843 billion, or $7,700 per household in the United States.[6]

3.3 Deregulating the Economy

Concern that regulation might sometimes fail to serve the public interest prompted a push to deregulate some industries, beginning in the late 1970s. In 1978, for example, Congress passed the Airline Deregulation Act, which removed many of the regulations that had prevented competition in the airline industry. Safety regulations were not affected. The results of deregulation included a substantial reduction in airfares, the merger and consolidation of airlines, and the emergence of frequent flier plans and other marketing schemes designed to increase passenger miles. Not all the consequences of deregulation were applauded, however. Many airlines, unused to the demands of a competitive, unprotected, and unregulated environment, went bankrupt or were taken over by other airlines. Large airlines abandoned service to small and midsized cities, and although most of these routes were picked up by smaller regional airlines, some consumers complained about inadequate service. Nevertheless, the more competitive airline system today is probably an improvement over the highly regulated industry that existed in the 1970s. It is certainly cheaper. Table 16.3 suggests that the improvements in consumer welfare from deregulation through the 1990s have been quite substantial across a broad spectrum of industries that have been deregulated.

TABLE 16.3 Improvement in Consumer Welfare from Deregulation

Economist Clifford Winston found substantial benefits from deregulation in the five industries he studied—airlines, trucking, railroads, banking, and natural gas.

Industry	Improvements
Airlines	Average fares are roughly 33% lower in real terms since deregulation, and service frequently has improved significantly.
Less-than-truckload trucking	Average rates per vehicle mile have declined at least 35% in real terms since deregulation, and service times have improved significantly.
Truckload trucking	Average rates per vehicle mile have declined by at least 75% in real terms since deregulation, and service times have improved significantly.
Railroads	Average rates per ton-mile have declined more than 50% in real terms since deregulation, and average transit time has fallen more than 20%.
Banking	Consumers have benefited from higher interest rates on deposits, from better opportunities to manage risk, and from more banking offices and automated teller machines.
Natural gas	Average prices for residential customers have declined at least 30% in real terms since deregulation, and average prices for commercial and industrial customers have declined more than 30%. In addition, service has been more reliable as shortages have been almost completely eliminated.

Source: Clifford Winston, "U.S. Industry Adjustment to Economic Deregulation," Journal of Economic Perspectives 12(3) (Summer 1998): 89–110.

But there are forces working in the opposite direction as well. Many businesses continue to turn to the government for protection from competition. Public choice theory suggests that more, not less, regulation is often demanded by firms threatened with competition at home and abroad. More and more people seem to demand environmental protection, including clear air, clean water, and regulation of hazardous waste and toxic waste. Indeed, as incomes rise over time, there is evidence that the demand for safety rises. This market phenomenon began before the birth of regulatory agencies and can be seen in the decline in unintentional injury deaths over the course of the last hundred years.[7] And there is little reason to expect less demand for regulations in the areas of civil rights, handicapped rights, gay rights, medical care, and elderly care.

The basic test of rationality—that marginal benefits exceed marginal costs—should guide the formulation of regulations. While economists often disagree about which, if any, consumer protection regulations are warranted, they do tend to agree that market incentives ought to be used when appropriate and that the most cost-effective policies should be pursued.

KEY TAKEAWAYS

- Federal, state, and local governments regulate the activities of firms and consumers.
- The public interest theory of regulation asserts that regulatory efforts act to move markets closer to their efficient solutions.
- The public choice theory of regulation argues that regulatory efforts serve private interests, not the public interest.
- Consumer protection efforts may sometimes be useful, but they tend to produce behavioral responses that often negate the effort at protection.
- Deregulation efforts through the 1990s generally produced large gains in consumer welfare, though demand for more regulation is rising in certain areas, especially finance.

TRY IT!

The deregulation of the airline industry has generally led to lower fares and increased quantities produced. Use the model of demand and supply to show this change. What has happened to consumer surplus in the market? (*Hint:* You may want to refer back to the earlier discussion of consumer surplus.)

Case in Point: Do Consumer Protection Laws Protect Consumers?

© 2010 Jupiterimages Corporation

Economist W. Kip Viscusi of the Harvard Law School has long advocated economic approaches to health and safety regulations. Economic approaches recognize 1) behavioral responses to technological regulations; 2) performance-oriented standards as opposed to command-and-control regulations; and 3) the opportunity cost of regulations. Below are some examples of how these economic approaches would improve health and safety policy.

Behavioral responses: Consider the requirement, imposed in 1972, that aspirin containers have childproof caps. That technological change seemed straightforward enough. But, according to Mr. Viscusi, the result has not been what regulators expected. Mr. Viscusi points out that childproof caps are more difficult to open. They thus increase the cost of closing the containers properly. An increase in the cost of any activity reduces the quantity demanded. So, childproof caps result in fewer properly closed aspirin containers.

Mr. Viscusi calls the response to childproof caps a "lulling effect." Parents apparently think of containers as safer and are, as a result, less careful with them. Aspirin containers, as well as other drugs with childproof caps, tend to be left open. Mr. Viscusi says that the tragic result is a dramatic increase in the number of children poisoned each year. Hence, he urges government regulators to take behavioral responses into account when promulgating technological solutions. He also advocates well-articulated hazard warnings that give consumers information on which to make their own choices.

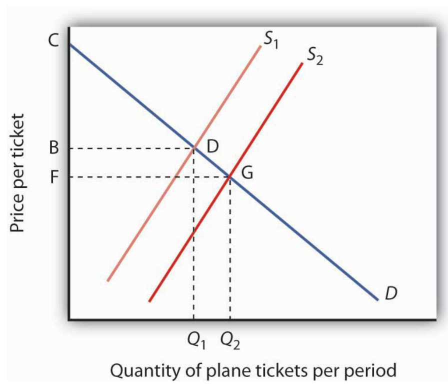
Performance-oriented standards: Once a health and safety problem has been identified, the economic approach would be to allow individuals or firms discretion in how to address the problem as opposed to mandating a precise solution. Flexibility allows a standard to be met in a less costly way and can have greater impact than command-and-control approaches. Mr. Viscusi cites empirical evidence that worker fatality rates would be about one-third higher were it not for the financial incentives firms derive from creating a safer workplace and thereby reducing the workers' compensation premiums they pay. In contrast, empirical estimates of the impact of OSHA regulations, most of which are of the command-and-control type, range from nil to a five to six percent reduction in worker accidents that involve days lost from work.

Opportunity cost of regulations: Mr. Viscusi has estimated the cost per life saved on scores of regulations. Some health and safety standards have fairly low cost per life saved. For example, car seat belts and airplane cabin fire protection cost about \$100,000 per life saved. Automobile side impact standards and the children's sleepwear flammability ban, at about \$1 million per life saved, are also fairly inexpensive. In contrast, the asbestos ban costs \$131 million per life saved, regulations concerning municipal solid waste landfills cost about \$23 billion per life saved, and regulations on Atrazine/alachlor in drinking water cost a whopping \$100 billion per life saved. "A regulatory system based on sound economic principles would reallocate resources from the high-cost to the low-cost regulations. That would result in more lives saved at the same cost to society."

Sources: W. Kip Viscusi, "Safety at Any Price?" Regulation, Fall 2002: 54–63; W. Kip Viscusi, "The Lulling Effect: The Impact of Protective Bottlecaps on Aspirin and Analgesic Poisonings," American Economic Review 74(2) (1984): 324–27.

ANSWER TO TRY IT! PROBLEM

Deregulation of the airline industry led to sharply reduced fares and expanded output, suggesting that supply increased. That should significantly increase consumer surplus. Specifically, the supply curve shifted from S_1 to S_2. Consumer surplus is the difference between the total benefit received by consumers and total expenditures by consumers. Before deregulation, when the price was B and the quantity was Q_1, the consumer surplus was BCD. The lower rates following deregulation reduced the price to consumers to, say, F, and increased the quantity to Q_2 on the graph, thereby increasing consumer surplus to FCG.



4. REVIEW AND PRACTICE

Summary

This chapter has shown that government intervention in markets takes the form of antitrust action to prevent the abuse of market power and regulations aimed at achieving socially desired objectives that are not or cannot be provided by an unregulated market system.

We saw that antitrust policy has evolved from a view that big business was bad business to an attempt to assess how the behavior of firms and the structure of markets affect social welfare and the public interest. The rule of reason rather than per se illegality guides most antitrust policy today, but because there is considerable debate concerning the appropriate role of government antitrust policy and regulation, application of the rule of reason in specific cases is uneven. Prosecution and enforcement of the nation's antitrust laws has varied over time.

The rising role of a global economy in the last half of the twentieth century reduced the degree of market power held by domestic firms. Policymakers have reconsidered antitrust policy and what types of joint ventures and cooperation among competing firms should be allowed. U.S. antitrust policy has not been abandoned, but since the early 1980s it has been applied with greater consideration of its implications for the competitiveness of U.S. businesses against Asian, European, and other firms. Whether or not antitrust laws among nations will be made more compatible with each other is an issue for the future.

We saw that there are many different schools of thought concerning regulation. One group believes that regulation serves the public interest. Another believes that much current regulation protects regulated firms from competitive market forces and that the regulators are captured by the firms they are supposed to regulate. Yet another group points out that the regulators may do little more than serve their own interests, which include increasing the bureaucratic reach of their agencies.

Finally, the chapter looked at the complex issues surrounding consumer protection regulations. Consumer protection legislation has costs, borne by consumers and taxpayers. Economists are not in agreement concerning which, if any, consumer protection regulations are warranted. They do agree, however, that market incentives ought to be used when appropriate and that the most cost-effective policies should be pursued.

CONCEPT PROBLEMS

1. Apex Manufacturing charges Zenith Manufacturing with predatory pricing (that is, selling below cost). What do you think the antitrust authorities will want to consider before they determine whether to prosecute Zenith for unfair practices in restraint of trade?
2. Some states require firms to close on Sunday. What types of firms support these laws? Why? What types of firms do not support these laws? Why?
3. Individual taxis in New York, Chicago, and many other cities must have permits, but there are only a fixed number of permits. The permits are typically sold in the marketplace. Who benefits from such a regulation?
4. What do you predict is the impact on workers' wages of safety regulations in the workplace if the labor market is competitive?
5. Many states require barbers and beauticians to be licensed. Using the public interest theory of regulation as a base, what, if any, arguments could you make to support such a regulation? Do you think consumers gain from such regulations? Why not just allow anyone to open up a barber shop or beauty salon?
6. Suppose a landowner is required to refrain from developing his or her land in order to preserve the habitat of an endangered species. The restriction reduces the value of the land by 50%, to $1 million. Under present law, the landowner does not have to be compensated. Several proposals considered by Congress would require that this landowner be compensated. How does this affect the cost of the regulation?
7. A study by the Federal Trade Commission compared the prices of legal services in cities that allowed advertising by lawyers to prices of those same services in cities that did not. It found that the prices of simple wills with trust provisions were 11% higher in cities that did not allow advertising than they were in cities that did.[8] This, presumably, suggests the cost of such regulation. What might be the benefits? Do you think that such advertising should be restricted?
8. Economist W. Kip Viscusi, whose work was cited in the Case in Point, and Gerald Cavallo studied the effects of federal regulations requiring cigarette lighter safety mechanisms.[9] Explain how this technological improvement might improve safety and how it might reduce safety.
9. Explain how licensing requirements for providers of particular services result in higher prices for such services. Are such requirements justified? Why or why not?
10. What is so bad about price-fixing? Why does the government prohibit it?
11. In a 1956 antitrust case against DuPont, the Justice Department argued that the firm held a near monopoly in the cellophane market. DuPont argued that the definition of the market should be changed to include all wrapping paper. Why is this issue of market definition important? (DuPont's position prevailed.)
12. The Case in Point on the efficacy of antitrust enforcement painted a rather negative view of antitrust enforcement. Do you agree with this assessment? Why or why not?
13. The Case in Point on Boeing and the European Union discussed a situation in which a foreign government, the European Union, attempted to exert authority over a relationship between two U.S. firms. How is this possible?

NUMERICAL PROBLEMS

1. In 1986, Pepsi announced its intention to buy 7-Up, and Coca-Cola proposed buying Dr Pepper. The table below shows the market shares held by the soft-drink firms in 1986. Assume that the remaining 15% of the market is composed of 15 firms, each with a market share of 1%.

Company	Market share (percent)
Coca-Cola	39
PepsiCo	28
Dr Pepper	7
7-Up	6

   1. Calculate the Herfindahl–Hirschman Index (HHI) for the industry as it was structured in 1986.
   2. Calculate the postmerger HHI if only PepsiCo had bought 7-Up.
   3. Calculate the postmerger HHI if only Coca-Cola had bought Dr Pepper.
   4. How would you expect the Justice Department to respond to each merger considered separately? To both?

   (By the way, the proposed mergers *were* challenged, and neither was completed.)

ENDNOTES

1. Eric Dash, "MasterCard Will Pay $1.8 Billion To a Rival," *New York Times*, June 26, 2008, p. C4; and United States vs. Visa U.S.A., Inc., 344 F.3d 229 (2d. Circuit 2003).
2. G. J. Stigler, "The Economists and the Problem of Monopoly," *American Economic Review Papers and Proceedings* 72:2 (May 1982): 8–9.
3. Matt Krantz, "Merger Market Arrives At 'Good spot'; 2006 the Busiest Takeover Year Since End Of '90s Bull," *USA Today*, November 7, 2006, p. 3B; and Matt Krantz, "Big Day for Buyouts, But Tepid Pace Forecase To Continue; Credit Crunch and Other Economic Fears Take Toll," *USA Today*, December 18, 2007, p. 1B.
4. Sam Peltzman, "The Effects of Automobile Safety Regulations," *Journal of Political Economy* 83 (August 1975): 677–725.
5. Alma Cohen and Liran Einan, "The Effects of Mandatory Seat Belt Laws on Driving Behaviour and Traffic Fatalities," *Review of Economics and Statistics* 85:4 (November 2003): 828–43.
6. W. Mark Crain and Thomas D. Hopkins, "The Impact of Regulatory Costs on Small Firms," Report for the Office of Advocacy, U.S. Small Business Administration, Washington, D.C., RFP No. SBAHQ-00-R-0027, October 2001, p. 1.
7. W. Kip Viscusi, "Safety at Any Price?" *Regulation*, Fall 2002: 54–63.
8. See Carolyn Cox and Susan Foster, "The Costs and Benefits of Occupational Regulation," Federal Trade Commission, October 1990, p. 31.
9. W. Kip Viscusi, "The Lulling Effect: The Impact of Protective Bottlecaps on Aspirin and Analgesic Poisonings," *American Economic Review* 74(2) (1984): 324–27.

CHAPTER 17

International Trade

START UP: TRADE WINDS

Rapid increases in the flow of goods and services between vastly different nations and cultures have changed what people eat, how they dress, and even how they communicate with one another. For you, increased trade has meant greater choice of what to buy and often lower prices.

Look through your room. Chances are it is full of items from all around the world. The relatively free trade that exists today provides you with expanded choices. No one forced you to buy that shirt from India or that CD player from Japan. Presumably you bought them because you preferred them to other shirts and CD players you might have bought, perhaps because they had certain characteristics—style, color, perceived quality, or price—that you favored.

Your gains are being experienced worldwide because the winds of international trade have blown generally freer in the past decades. Nations all over the world have dramatically lowered the barriers they impose on the products of other countries.

One region that was once closed to virtually all trade but is now open is Eastern Europe and the countries that made up the former Soviet Union. A key part of these countries' attempts to create market capitalist economic systems has been the opening of their borders to international trade.

In Western Europe, the members of the European Union (EU) have eliminated virtually every restriction on the free flow of goods and services among them. A truckload of electronic equipment from Italy now passes through France on its way to Spain with no more restrictions than would be encountered by a truck delivering goods from Michigan to Illinois. The purchase of the equipment can even be arranged using a new currency, the euro, which has been adopted by most EU nations.

Canada, Mexico, and the United States, while not adopting a common currency, have created a similar free trade area, the North American Free Trade Area (NAFTA). In addition, the 18 member nations of the Asian-Pacific Economic Cooperation organization (APEC) agreed in 1994 to forge a free trade area among industrialized nations such as the United States and Japan by 2010. Other member nations such as Mexico and China agreed to participate by 2020.

NAFTA has resulted in a dramatic increase in trade between Canada, the United States, and Mexico. Since NAFTA's creation in 1994, employment in all three countries has risen substantially. Those increases in employment cannot necessarily be attributed to NAFTA, but the fact that they have occurred flies in the face of the arguments when NAFTA was first proposed that it would lead to a reduction in U.S. employment.

President Bush proposed and Congress passed in 2005 the creation of a Central American Free Trade Association (CAFTA) that would create a free trade area south of Mexico and linked to the United States. It abolishes most tariff restrictions between the United States and the countries of Central America—Costa Rica, the Dominican Republic, El Salvador, Guatemala, Honduras, and Nicaragua. The six countries make up the second-largest export market for the United States in Latin America, behind Mexico. President Bush has also proposed extending the free trade zone throughout the Western Hemisphere.

And, in 1995, the World Trade Organization (WTO) was established to "help trade flow smoothly, freely, fairly and predictably" among member nations. In 2008, it had 153 member countries. Since World War II, the General Agreement on Tariffs and Trade (GATT)—WTO's predecessor—and WTO have generated a series of agreements that slashed trade restraints among members. These agreements have helped propel international trade, which in 2006 was more than 35 times its level in 1950, but the negotiations leading to these agreements have always been protracted and tumultuous and issues of nationalism and patriotism are often not far from the surface. The current and ninth round of trade talks are referred to as the Doha Round, because they were officially launched in Doha, Qatar, in 2001. In mid-2008, talks were still mired in controversy over the removal of agricultural export subsidies and lowering of trade barriers of various kinds.

Why have so many countries moved to make trade freer? What are the effects of free trade? Why do efforts to eliminate trade restrictions meet with resistance? Why do many nations continue to impose barriers against some foreign goods and services? How do such barriers affect the economy? How do such barriers affect *you?*

This chapter will answer these questions by developing a model of international trade based on the idea of comparative advantage, introduced in an earlier chapter. The model predicts that free international trade will benefit the countries that participate in it. Free trade does not benefit every individual, however. Most people benefit from free trade, but some are hurt. We will then look at the phenomenon of two-way trade, in which countries both import and export the same goods. The last part of the chapter examines the effects of trade restrictions and evaluates the arguments made for such restrictions. Economists tend to be skeptical of their validity.

1. THE GAINS FROM TRADE

LEARNING OBJECTIVES

1. **Differentiate between an absolute advantage in producing some good and a comparative advantage.**
2. **Explain and illustrate the conditions under which two countries can mutually benefit from trading with each other.**
3. **Explain and illustrate how the terms of trade determine the extent to which each country specializes.**
4. **Explain and illustrate the mutual benefits of trade.**

To model the effects of trade, we begin by looking at a hypothetical country that does not engage in trade and then see how its production and consumption change when it does engage in trade.

1.1 Production and Consumption Without International Trade

Suppose the hypothetical country of Roadway is completely isolated from the rest of the world. It neither exports nor imports goods and services. We shall use the production possibilities model to analyze Roadway's ability to produce goods and services.

A production possibilities curve illustrates the production choices available to an economy. Recall that the production possibilities curve for a particular country is determined by the factors of production and the technology available to it.

Figure 17.1 shows a production possibilities curve for Roadway. We assume that it produces only two goods—trucks and boats. Roadway must be operating somewhere on its production possibilities curve or it will be wasting resources or engaging in inefficient production. If it were operating inside the curve at a point such as D, then a combination on the curve, such as B, would provide more of both goods (Roadway produces 3,000 more trucks and 3,000 more boats per year at B than at D). At any point inside the curve, Roadway's production would not be efficient. Point E suggests an even higher level of output than points A, B, or C, but because point E lies outside Roadway's production possibilities curve, it cannot be attained.

We have learned that the absolute value of the slope of a production possibilities curve at any point gives the quantity of the good on the vertical axis that must be given up to produce an additional unit of the good on the horizontal axis. It thus gives the opportunity cost of producing another unit of the good on the horizontal axis.

FIGURE 17.1 Roadway's Production Possibilities Curve

The production possibilities curve for Roadway shows the combinations of trucks and boats that it can produce, given the factors of production and technology available to it. To maximize the value of total production, Roadway must be operating somewhere along this curve. Production at point D implies that Roadway is failing to use its resources fully and efficiently; production at point E is unobtainable.

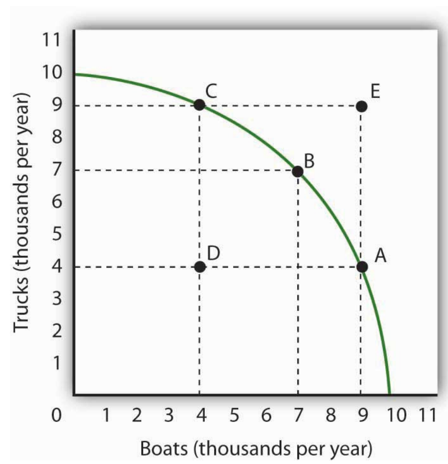


FIGURE 17.2 Measuring Opportunity Cost in Roadway

The slope of the production possibilities curve at any point is equal to the slope of a line tangent to the curve at that point. The absolute value of the slope equals the opportunity cost of increased boat production. Moving down and to the right along its production possibilities curve, the opportunity cost of boat production increases; this is an application of the law of increasing opportunity cost.

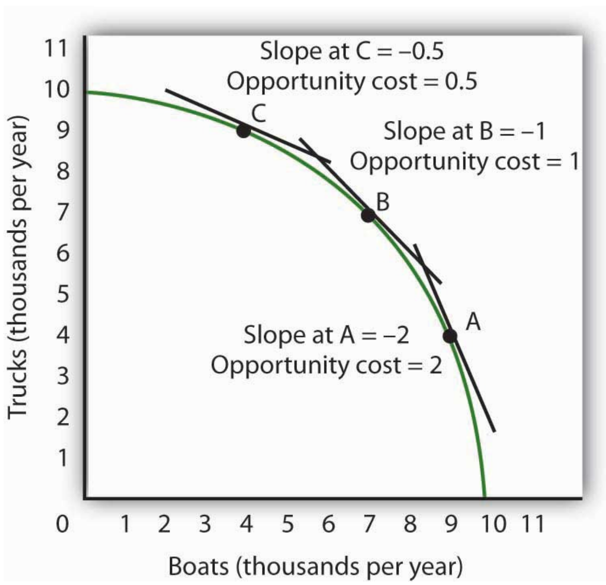


Figure 17.2 shows the opportunity cost of producing boats at points A, B, and C. Recall that the slope of a curve at any point is equal to the slope of a line drawn tangent to the curve at that point. The slope of a line tangent to the production possibilities curve at point B, for example, is −1. The opportunity cost of producing one more boat is thus one truck. As the law of increasing opportunity costs predicts, in order to produce more boats, Roadway must give up more and more trucks for each additional boat. Roadway's opportunity cost of producing boats increases as we travel down and to the right on its production possibilities curve.

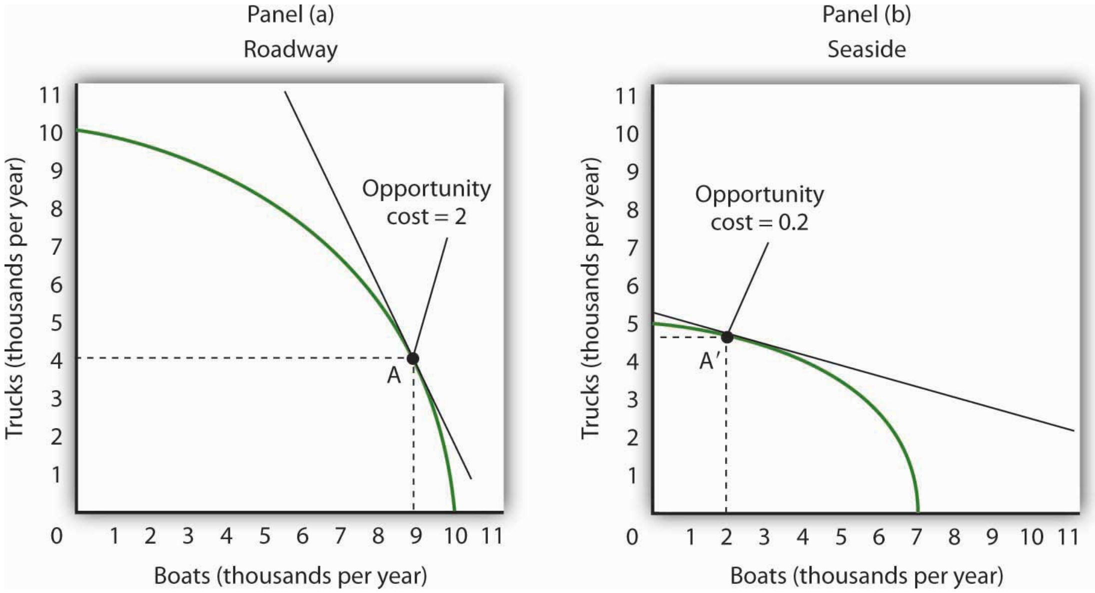
1.2 Comparative Advantage

People participate in international trade because they make themselves better off by doing so. In this section we will find that countries that participate in international trade are able to consume more of all goods and services than they could consume while producing in isolation from the rest of the world.

Suppose the world consists of two countries, Roadway and Seaside. Their production possibilities curves are given in Figure 17.3. Roadway's production possibilities curve in Panel (a) is the same as the one in Figure 17.1 and Figure 17.2. Seaside's curve is given in Panel (b).

FIGURE 17.3 Comparative Advantage in Roadway and Seaside

Roadway's production possibilities curve is given in Panel (a); it is the same one we saw in Figure 17.1 and Figure 17.2. The production possibilities curve for a second hypothetical country, Seaside, is given in Panel (b). If no trade occurs between the two countries, suppose that Roadway is at Point A and that Seaside is at Point A′. Notice that the opportunity cost of an additional boat in Roadway is two trucks, while the opportunity cost of an additional boat in Seaside is 0.2 trucks. Clearly, Seaside has a comparative advantage in the production of boats.



Each country produces two goods, boats and trucks. Suppose no trade occurs between the two countries and that they are each currently operating on their production possibilities curves at points A and A′ in Figure 17.3. We will assume that the two countries have chosen to operate at these points through the workings of demand and supply. That is, resources have been guided to their current uses as producers have responded to the demands of consumers in the two countries. In turn, consumers have responded to the prices charged by sellers of boats and trucks.

absolute advantage

Situation in which an economy or individual can produce more of any good per unit of labor than another country or individual.

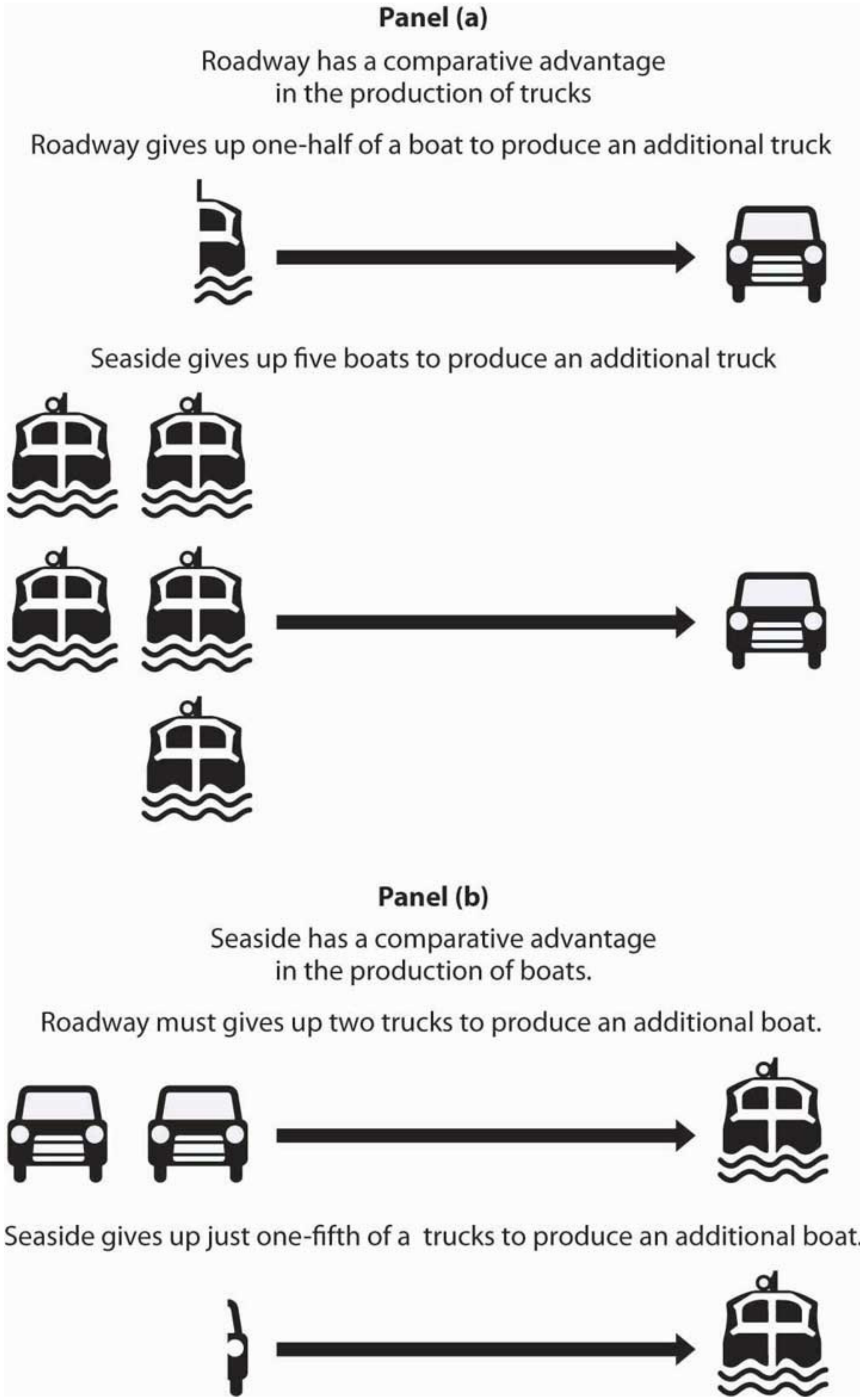
The two countries differ in their respective abilities to produce trucks and boats. As we can see by looking at the intersection of the production possibilities curves with the vertical axes in Figure 17.3, Roadway is able to produce more trucks than Seaside. If Roadway concentrated all of its resources on the production of trucks, it could produce 10,000 trucks per year. Seaside could produce only 5,000. Now look at the intersection of the production possibilities curves with the horizontal axes. If Roadway concentrated all of its resources on the production of boats, it could produce 10,000 boats. Seaside could produce only 7,000 boats. Because Roadway is capable of producing more of both goods, we can infer that it has more resources or is able to use its labor and capital resources more productively than Seaside. When an economy or individual can produce more of any good per unit of labor than another country or individual, that country or person is said to have an **absolute advantage.**

Despite the fact that Roadway can produce more of both goods, it can still gain from trade with Seaside—and Seaside can gain from trade with Roadway. The key lies in the opportunity costs of the two goods in the two countries. The country with a lower opportunity cost for a particular good or service has a comparative advantage in producing it and will export it to the other country.

We can determine opportunity costs in the two countries by comparing the slopes of their respective production possibilities curves at the points where they are producing. At point A in Panel (a) of Figure 17.3, one additional boat costs two trucks in Roadway; that is its opportunity cost. At point A′ in Panel (b), 1 additional boat in Seaside costs only 0.2 truck. Alternatively, we can ask about the opportunity cost of an additional truck. In Roadway, an additional truck costs 0.5 boats. In Seaside, it costs five boats. Roadway thus has a comparative advantage in producing trucks; Seaside has a comparative advantage in producing boats. This situation is suggested pictorially in Figure 17.4.

FIGURE 17.4 A Picture of Comparative Advantage in Roadway and Seaside

The exhibit gives a picture of Roadway's comparative advantage in trucks and Seaside's comparative advantage in boats.



1.3 Specialization and the Gains from Trade

We have so far assumed that no trade occurs between Roadway and Seaside. Now let us assume that trade opens up. The fact that the opportunity costs differ between the two countries suggests the possibility for mutually advantageous trade. The opportunities created by trade will induce a greater degree of specialization in both countries, specialization that reflects comparative advantage.

Trade and Specialization

Before trade, truck producers in Roadway could exchange a truck for half a boat. In Seaside, however, a truck could be exchanged for five boats. Once trade opens between the two countries, truck producers in Roadway will rush to export trucks to Seaside.

Boat producers in Seaside enjoy a similar bonanza. Before trade, one of their boats could be exchanged for one-fifth of a truck. By shipping their boats to Roadway, they can get two trucks for each boat. Boat producers in Seaside will rush to export boats to Roadway.

terms of trade

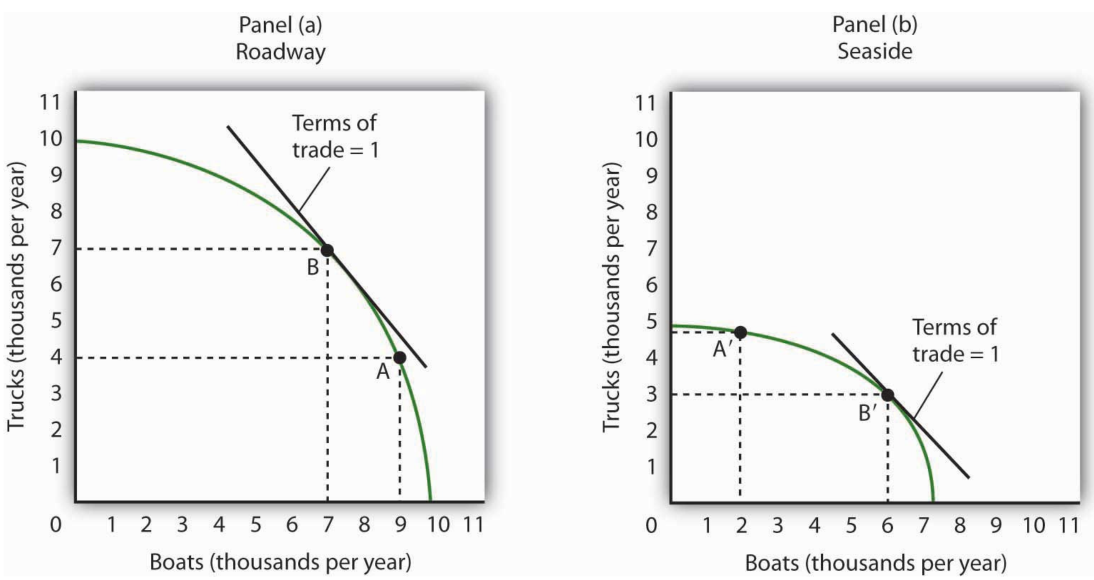
The rate at which a country can trade domestic products for imported products.

Once trade between Roadway and Seaside begins, the **terms of trade**, the rate at which a country can trade domestic products for imported products, will seek market equilibrium. The final terms of trade will be somewhere between one-half boats for one truck found in Roadway and five boats for one truck in Seaside. Suppose the terms of trade are one boat for one truck. (How the specific terms of trade are actually determined is not important for this discussion. It is enough to know that the final terms of trade will lie somewhere between Seaside's and Roadway's opportunity costs for boat and truck production.) Roadway's truck producers will now get one boat per truck—a far better exchange than was available to them before trade.

Roadway's manufacturers will move to produce more trucks and fewer boats until they reach the point on their production possibilities curve at which the terms of trade equals the opportunity cost of producing trucks. That occurs at point B in Panel (a) of Figure 17.5; Roadway now produces 7,000 trucks and 7,000 boats per year.

FIGURE 17.5 International Trade Induces Greater Specialization

Before trade, Roadway is producing at point A in Panel (a) and Seaside is producing at point A′ in Panel (b). The terms of trade are one, meaning that one boat exchanges for one truck. Roadside moves along its production possibilities curve to point B, at which the curve has a slope of −1. Roadside will produce more trucks (and fewer boats). Seaside moves along its production possibilities curve to point B′, at which the slope equals −1. Seaside will produce more boats (and fewer trucks). Trade leads each country in the direction of producing more of the good in which it has a comparative advantage.



Similarly, Seaside will specialize more in boat production. As shown in Panel (b) of Figure 17.5, producers will shift resources out of truck production and into boat production until they reach the point on their production possibilities curve at which the terms of trade equal the opportunity cost of producing boats. This occurs at point B′; Seaside produces 3,000 trucks and 6,000 boats per year.

We see that trade between the two countries causes each country to specialize in the good in which it has a comparative advantage. Roadway produces more trucks, and Seaside produces more boats. The specialization is not, however, complete. The law of increasing opportunity cost means that, as an economy moves along its production possibilities curve, the cost of additional units rises. An economy with a comparative advantage in a particular good will expand its production of that good only up to the point where its opportunity cost equals the terms of trade.

As a result of trade, Roadway now produces more trucks and fewer boats. Seaside produces more boats and fewer trucks. Through exchange, however, both countries are likely to end up consuming more of *both* goods.

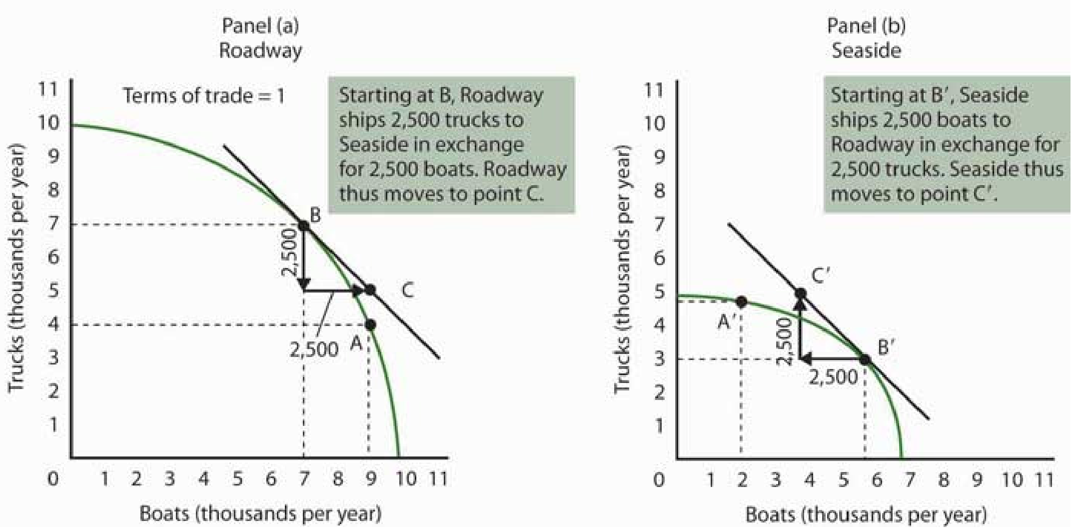
Figure 17.6 shows one such possibility. Suppose Roadway ships 2,500 trucks per year to Seaside in exchange for 2,500 boats, as shown in the table in Figure 17.6. Roadway thus emerges with 4,500 trucks (the 7,000 it produces at B minus the 2,500 it ships) and 9,500 boats. It has 500 more of each good than it did before trade. The precise amounts of each good shipped will depend on demand an supply. The essential point is that Roadway will produce more of the good—trucks—in which it has a comparative advantage. It will export that good to a country, or countries, that has a comparative advantage in something else.

FIGURE 17.6 The Mutual Benefits of Trade

Roadway and Seaside each consume more of both goods when there is trade between them. The table shows values of production before trade (BT) and after trade (AT). Here, the terms of trade are one truck in exchange for one boat. As shown in Panel (a) and in the exhibit's table, Roadway exports 2,500 trucks to Seaside in exchange for 2,500 boats and ends up consuming at point C, which is outside its production possibilities curve. Similarly, in Panel (b), Seaside ends up consuming at point C′, which is outside its production possibilities curve. Trade allows both countries to consume more than they are capable of producing.

Production and Consumption of Trucks and Boats

	Roadway				Seaside			
	Production		Consumption		Production		Consumption	
	BT	AT	BT	AT	BT	AT	BT	AT
Trucks	4,000	7,000	4,000	4,500	4,750	3,000	4,750	5,500
Boats	9,000	7,000	9,000	9,500	2,000	6,000	2,000	3,500



How does Seaside fare? When trade began, factors of production shifted into boat production, in which Seaside had a comparative advantage. Seaside tripled its production of boats—from 2,000 per year to 6,000 per year. It sends 2,500 of those boats to Roadway, so it ends up with 3,500 boats per year. It reduces its production of trucks to 3,000 per year, but receives 2,500 more from Roadway. That leaves it with 5,500. Seaside emerges from the opening of trade with 1,500 more boats and 750 more trucks than it had before trade.

As Roadway trades trucks for boats, its production remains at point B. But it now consumes combination C; it has more of both goods than it had at A, the solution before trade. Seaside's production remains at point B′, but it now consumes at point C′, where it has more trucks and more boats than it had before trade.

Although all countries can increase their consumption through trade, not everyone in those countries will be happy with the result. In the case of Roadway and Seaside, for example, some boat producers in Roadway will be displaced as cheaper boats arrive from Seaside. Some truck producers in Seaside will be displaced as cheaper trucks arrive from Roadway. The production possibilities model suggests that the resources displaced will ultimately find more productive uses. They will produce trucks in Roadway and boats in Seaside. *But* there will be a period of painful transition as workers and owners of capital and natural resources move from one activity to another. That transition will be completed when the two countries are back on their respective production possibilities curves. Full employment will be restored, which means both countries will be back at the same level of employment they had before trade.

Finally, note the fact that the two countries end up at C (Panel (a)) and C′ (Panel (b)). These points lie *outside* the production possibilities curves of both countries. Notice that each country *produces* on its production possibilities curve, but international trade allows both countries to *consume* a combination of goods they would be incapable of producing!

We see this same phenomenon in individual households. Each household specializes in an activity in which it has a comparative advantage. For one household, that may be landscaping, for another, it may be the practice of medicine, for another it may be the provision of childcare. Whatever the activity, specialization allows the household to earn income that can be used to purchase housing, food, clothing, and so on. Imagine for a moment how your household would fare if it had to produce every good or service it consumed. The members of such a household would work very hard, but it is inconceivable that the household could survive if it relied on itself for everything it consumed. By specializing in the activity in which each individual has a comparative advantage, people are able to consume far more than they could produce themselves.

Despite the transitional problems affecting some factors of production, the potential benefits from free trade are large. For this reason, most economists are strongly in favor of opening markets and extending international trade throughout the world. The economic case has been a powerful force in moving the world toward freer trade.

KEY TAKEAWAYS

- In order to maximize the value of its output, a country must be producing a combination of goods and services that lies on its production possibilities curve.
- Suppose two countries each produce two goods and their opportunity costs differ. If this is the case, there is an opportunity for trade between the two countries that will leave both better off.
- International trade leads countries to specialize in goods and services in which they have a comparative advantage.
- The terms of trade determine the extent to which each country will specialize. Each will increase production of the good or service in which it has a comparative advantage up to the point where the opportunity cost of producing it equals the terms of trade.
- Free international trade can increase the availability of all goods and services in all the countries that participate in it. Trade allows countries to consume combinations of goods and services they would be unable to produce.
- While free trade increases the total quantity of goods and services available to each country, there are both winners and losers in the short run.

TRY IT!

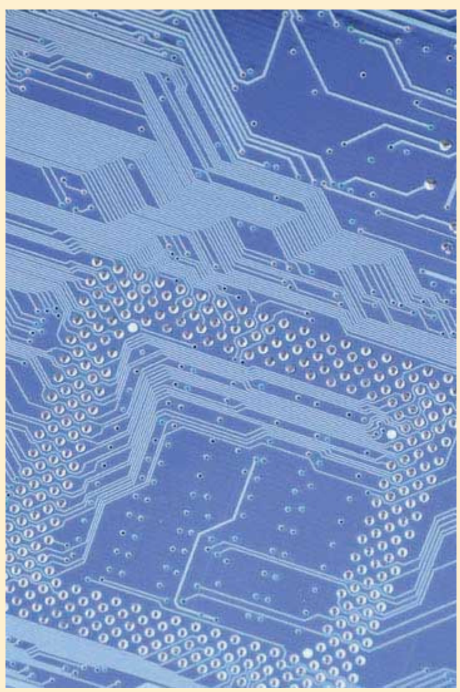
Suppose the world consists of two countries, Alpha and Beta. Both produce only two goods, computers and washing machines. Suppose that Beta is much more populous than Alpha, but because workers in Alpha have more physical and human capital, Alpha is able to produce more of both goods than Beta.

Specifically, suppose that if Alpha devotes all its factors of production to computers, it is able to produce 10,000 per month, and if it devotes all its factors of production to washing machines, it is able to produce 10,000 per month. Suppose the equivalent amounts for Beta are 8,000 computers and 8,000 washing machines per month. Sketch typical, bowed-out production possibilities curves for the two countries. (You only have numbers for the end points of the production possibilities curves. Use them to sketch curves of a typical shape. Place washing machines on the vertical axis and computers on the horizontal axis.)

Assume the computers and washing machines produced in the two countries are identical. Assume that no trade occurs between the two countries. In Alpha, at the point on its production possibilities curve at which it is operating, the opportunity cost of an additional washing machine is 0.5 computers. At the point on its production possibilities curve at which it is operating, the opportunity cost of an additional washing machine in Beta is 3.5 computers. How many computers exchange for a washing machine in Alpha? Beta?

Now suppose trade occurs, and the terms of trade are two washing machines for one computer. How will the production of the two goods be affected in each economy? Show your results graphically and explain them.

Case in Point: The U.S. Comparative Advantage in High-Tech Capital Goods and Services



© 2010 Jupiterimages Corporation

A flight across the United States almost gives a birds-eye view of an apparent comparative advantage for the United States. One sees vast expanses of farmland. Surely agricultural goods represent an important comparative advantage for the United States.

Indeed, agricultural goods did once dominate American exports. Today, however, agricultural goods make up a small percentage of U.S. exports, though the amount of agricultural goods that the United States does export continues to grow.

Doomsayers suggest that our comparative advantage in the twenty-first century will lie in flipping hamburgers and sweeping the floors around Japanese computers. This forecast makes for good jokes, but it hardly squares with the facts. Recently America's comparative advantages lie in certain stages of the production process and in areas of the service sector.

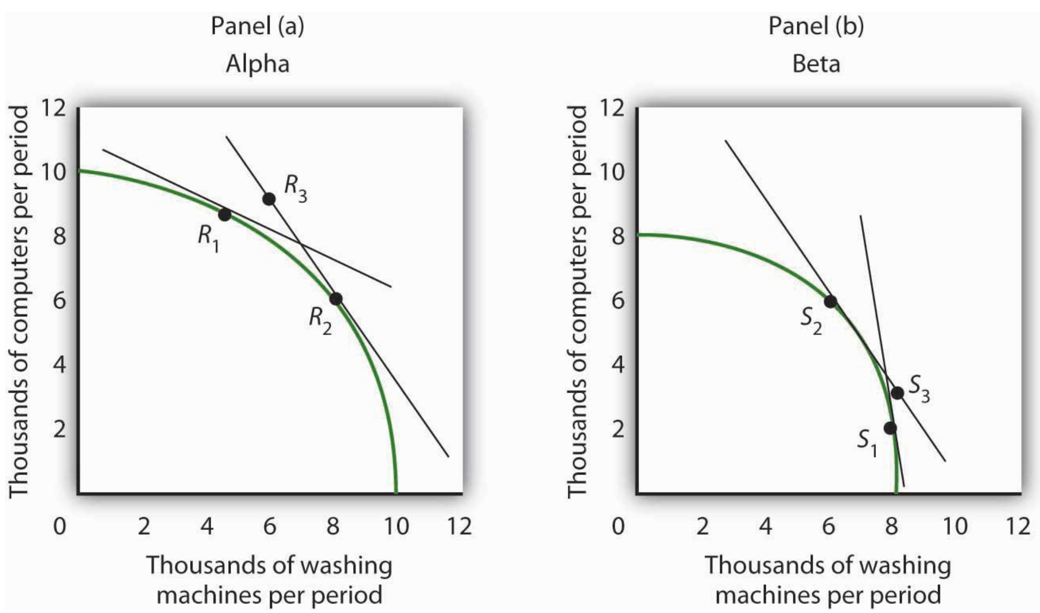
According to economist Catherine Mann of the Brookings Institution, "the United States has the comparative advantage in producing and exporting certain parts of the production process (the high-valued processor chips, the innovative and complex software, and the fully assembled product), but has relinquished parts of the production process to other countries where that stage of processing can be completed more cheaply (memory chips, 'canned' software, and most peripherals)."

In the area of services, Mann reports, the United States excels primarily in a rather obscure sounding area called "other private services," which, she contends, corresponds roughly to new economy services. Other private services include such areas as education, financial services, and business and professional services. This category of services has grown relentlessly over the past 15 years, despite cyclical downturns in other sectors. The United States developed its comparative advantage in these services as the share of services in the U.S. economy grew over time. She predicts that, as the economies of our trading partners grow, their demand for services will also increase. So, from a policy perspective, it is important for the U.S. to promote trading policies that will keep this sector open.

Sources: Catherine L. Mann, "Is the U.S. Trade Deficit Sustainable?" Washington, D.C: Brookings Institution, 1999; Catherine L. Mann, "The U.S. Current Account, New Economy Services, and Implications for Sustainability," Review of International Economics 12:2 (May 2004): 262–76.

ANSWER TO TRY IT! PROBLEM

Here are sketches of possible production possibilities curves. Alpha is operating at a point such as R_1, while Beta is operating at a point such as S_1. In Alpha, 1 computer trades for 2 washing machines; in Beta, 3.5 computers trade for one washing machine. If trade opens between the two economies and the terms of trade are 1.5, then Alpha will produce more washing machines and fewer computers (moving to a point such as R_2), while Beta will produce more computers and fewer washing machines (moving to a point such as S_2). Though you were not asked to do this, the graphs demonstrate that it is possible that trade will result in both countries having more of both goods. If, for example, Alpha ships 2,000 washing machines to Beta in exchange for 3,000 computers, then the two economies will move to points R_3 and S_3, respectively, consuming more of both goods than they had before trade. There are many points along the tangent lines drawn at points R_2 and S_2 that are up to the right and therefore contain more of both goods. We have chosen points R_3 and S_3 at specific points, but any point along the tangent line that is up to the right from R_1 and S_1 would suffice to illustrate the fact that both countries can end up consuming more of both goods.



2. TWO-WAY TRADE

LEARNING OBJECTIVES

1. Distinguish between one-way trade and two-way trade.
2. Explain why two-way trade may occur.

The model of trade presented thus far assumed that countries specialize in producing the good in which they have a comparative advantage and, therefore, engage in one-way trade. **One-way (or interindustry) trade** occurs when countries specialize in producing the goods in which they have a comparative advantage and then export those goods so they can import the goods in which they do not have a comparative advantage.

one-way (or interindustry) trade

Situation in which countries specialize in producing the goods in which they have a comparative advantage and then export those goods so they can import the goods in which they do not have a comparative advantage.

However, when we look at world trade, we also see countries exchanging the same goods or goods in the same industry category. For example, the United States may both export construction materials to Canada and import them from Canada. American car buyers can choose Chevrolets, Fords, and Chryslers. They can also choose imported cars such as Toyotas. Japanese car buyers may choose to purchase Toyotas—or imported cars such as Chevrolets, Fords, and Chryslers. The United States imports cars from Japan and exports cars to it. Conversely, Japan imports cars from the United States and exports cars to it. International trade in which countries both import and export the same or similar goods is called **two-way (or intraindustry) trade.**

two-way (or intraindustry) trade

International trade in which countries both import and export the same or similar goods.

Two reasons countries import and export the same goods are variations in transportation costs and seasonal effects. In the example of the United States and Canada both importing and exporting construction materials, transportation costs are the likely explanation. It may be cheaper for a contractor in northern Maine to import construction materials from the eastern part of Canada than to buy them in the United States. For a contractor in Vancouver, British Columbia, it may be cheaper to import construction materials from somewhere in the western part of the United States than to buy them in Canada. By engaging in trade, both the American and Canadian contractors save on transportation costs. Seasonal factors explain why the United States both imports fruit from and exports fruit to Chile.

Another explanation of two-way trade in similar goods lies in recognizing that not all goods are produced under conditions of perfect competition. Once this assumption is relaxed, we can explain two-way trade in terms of a key feature of monopolistic competition and some cases of oligopoly: product differentiation. Suppose two countries have similar endowments of factors of production and technologies available to them, but their products are differentiated—clocks produced by different manufacturers, for example, are different. Consumers in the United States buy some clocks produced in Switzerland, just as consumers in Switzerland purchase some clocks produced in the United States. Indeed, if two countries are similar in their relative endowments of factors of production and in the technologies available to them, two-way trade based on product differentiation is likely to be more significant than one-way trade based on comparative advantage.

In comparison to the expansion of one-way trade based on comparative advantage, expansion of two-way trade may entail lower adjustment costs. In the case of two-way trade, there is specialization within industries rather than movement of factors of production out of industries that compete with newly imported goods and into export industries. Such adjustments are likely to be faster and less painful for labor and for the owners of the capital and natural resources involved.

Because two-way trade often occurs in the context of imperfect competition, we cannot expect it to meet the efficiency standards of one-way trade based on comparative advantage and the underlying assumption of perfectly competitive markets. But, as we discussed in the chapter on imperfect competition, the inefficiency must be weighed against the benefits derived from product differentiation. People in the United States are not limited to buying only the kinds of cars produced in the United States, just as people in Japan are not limited to buying only cars produced in Japan.

KEY TAKEAWAYS

- Specialization and trade according to comparative advantage leads to one-way trade.
- A large percentage of trade among countries with similar factor endowments is two-way trade, in which countries import and export the same or similar goods and services.
- Two-way trade is often explained by variations in transportation costs and seasonal factors; in similar goods it often occurs in the context of models of imperfect competition.
- Adjustment costs associated with expansion of two-way trade may be lower than for expansion of one-way trade.

TRY IT!

The text argues that two-way trade must be a result of transportation cost, climate, or imperfect competition. Explain why.

Case in Point: Two- Way Trade in Water: A Growth Industry

In the 1930s, the successful introduction into the United States of French-made Perrier showed that U.S. consumers were open to a "new" bottled beverage. Since then, the U.S. bottled water business has taken off and bottled water is now the second largest commercial beverage category by volume, after carbonated soft drinks.

Seeing the increased popularity of bottled water, both PepsiCo and Coca-Cola launched their own bottled water brands, Aquafina and Dasani, respectively. Both of these brands are made from purified tap water. Dasani has minerals added back into it; Aquafina does not. Other brands of water come from springs or artesian wells. While domestic brands of water have multiplied, Americans still drink some imported brands. Representing only about 3% of the U.S. market, the volume of imported water nonetheless has about doubled in the last five years.

And U.S. bottled water companies are also eyeing markets in other countries. As *New York Times* columnist and book author Thomas Friedman noted as he was being shown around a customer call center in Bangalore, India, the water on the desktops of the telemarketers was none other than Coke's Dasani.

Whether the differences in brands of water are perceived or real, it may not be too long before restaurants develop water lists next to their beer and wine lists. In the U.S. and in other countries around the world, there is likely to be a domestic section and an imported section on those lists. Two-way trade in water seems destined to be a growth industry for some time to come.

Sources: Thomas L. Friedman, "What Goes Around…" The New York Times, February 26, 2004, p. A27; Tom McGrath and Kate Dailey, "Liquid Assets," Men's Health 19:2 (March 2004): 142–49; Statistics from Beverage Marketing Corporation at www.bottledwater.org/public/Stats_2004.doc.

ANSWER TO TRY IT! PROBLEM

In the absence of one of these factors, there would only be one-way, or interindustry, trade, which would take place according to comparative advantage, as described in the first section of this chapter, with a country specializing in and exporting the goods in which it has a comparative advantage and importing goods in which it does not. Efficiency differences would be the only basis for trade.

3. RESTRICTIONS ON INTERNATIONAL TRADE

LEARNING OBJECTIVES

1. **Define the term protectionist policy and illustrate the general impact in a market subject to protectionist policy.**
2. **Describe the various forms of protectionist policy.**
3. **Discuss and assess the arguments used to justify trade restrictions.**

In spite of the strong theoretical case that can be made for free international trade, every country in the world has erected at least some barriers to trade. Trade restrictions are typically undertaken in an effort to protect companies and workers in the home economy from competition by foreign firms. A **protectionist policy** is one in which a country restricts the importation of goods and services produced in foreign countries. The slowdown in the U.S. economy late in 2007 and in 2008 has produced a new round of protectionist sentiment—one that became a factor in the 2008 U.S. presidential campaign.

protectionist policy

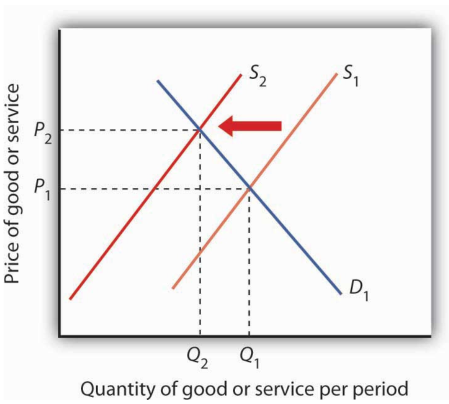
Policy that restricts the importation of goods and services produced in foreign countries.

The United States, for example, uses protectionist policies to limit the quantity of foreign-produced sugar coming into the United States. The effect of this policy is to reduce the supply of sugar in the U.S. market and increase the price of sugar in the United States. The 2008 U.S. Farm Bill sweetened things for sugar growers even more. It raised the price they are guaranteed to receive and limited imports of foreign sugar so that American growers will always have at least 85% of the domestic market. The bill for the first time set an income limit—only growers whose incomes fall below $1.5 million per year (for couples) or $750,000 for individuals will receive direct subsidies.[1]

The U.S. price of sugar is almost triple the world price of sugar, thus reducing the quantity consumed in the United States. The program benefits growers of sugar beets and sugar cane at the expense of consumers.

FIGURE 17.10 The Impact of Protectionist Policies

Protectionist policies reduce the quantities of foreign goods and services supplied to the country that imposes the restriction. As a result, such policies shift the supply curve to the left for the good or service whose imports are restricted. In the case shown, the supply curve shifts to S_2, the equilibrium price rises to P_2, and the equilibrium quantity falls to Q_2.



Source: Historical Statistics, Colonial Times to 1970: Statistical Abstract of the United States 1998, Table no. 1325; Statistical Abstract of the United States, 1990; U.S. International Commission (http://dataweb.usitc.gov/prepared_reports.asp).

In general, protectionist policies imposed for a particular good always reduce its supply, raise its price, and reduce the equilibrium quantity, as shown in Figure 17.11. Protection often takes the form of an import tax or a limit on the amount that can be imported, but it can also come in the form of voluntary export restrictions and other barriers.

3.1 Tariffs

A **tariff** is a tax on imported goods and services. The average tariff on dutiable imports in the United States (that is, those imports on which a tariff is imposed) is about 4%. Some imports have much higher tariffs. For example, the U.S. tariff on imported frozen orange juice is 35 cents per gallon (which amounts to about 40% of value). The tariff on imported canned tuna is 35%, and the tariff on imported shoes ranges between 2% and 48%.

A tariff raises the cost of selling imported goods. It thus shifts the supply curve for goods to the left, as in Figure 17.10. The price of the protected good rises and the quantity available to consumers falls.

3.2 Antidumping Proceedings

One of the most common protectionist measures now in use is the antidumping proceeding. A domestic firm, faced with competition by a foreign competitor, files charges with its government that the foreign firm is **dumping**, or charging an "unfair" price. Under rules spelled out in international negotiations that preceded approval of the World Trade Organization, an unfair price was defined as a price below production cost or below the price the foreign firm charges for the same good in its own country. While these definitions may seem straightforward enough, they have proven to be quite troublesome. The definition of "production cost" is a thoroughly arbitrary procedure. In defining cost, the government agency invariably includes a specification of a "normal" profit. That normal profit can be absurdly high. The United States Department of Justice, which is the U.S. agency in charge of determining whether a foreign firm has charged an unfair price, has sometimes defined normal profit rates as exceeding production cost by well over 50%, a rate far higher than exists in most U.S. industry.

tariff

A tax on imported goods and services.

dumping

The practice of a foreign firm charging a price in other countries that is below the price it charges in its home country.

The practice of a foreign firm charging a price in the United States that is below the price it charges in its home country is common. The U.S. market may be more competitive, or the foreign firm may simply be trying to make its product attractive to U.S. buyers that are not yet accustomed to its product. In any event, such price discrimination behavior is not unusual and is not necessarily "unfair."

In the United States, once the Department of Justice has determined that a foreign firm is guilty of charging an unfair price, the U.S. International Trade Commission must determine that the foreign firm has done material harm to the U.S. firm. If a U.S. firm has suffered a reduction in sales and thus in employment it will typically be found to have suffered material harm, and punitive duties will be imposed.

3.3 Quotas

A **quota** is a direct restriction on the total quantity of a good or service that may be imported during a specified period. Quotas restrict total supply and therefore increase the domestic price of the good or service on which they are imposed. Quotas generally specify that an exporting country's share of a domestic market may not exceed a certain limit.

quota

A direct restriction on the total quantity of a good or service that may be imported during a specified period.

In some cases, quotas are set to raise the domestic price to a particular level. Congress requires the Department of Agriculture, for example, to impose quotas on imported sugar to keep the wholesale price in the United States above 22 cents per pound. The world price is typically less than 10 cents per pound.

A quota restricting the quantity of a particular good imported into an economy shifts the supply curve to the left, as in Figure 17.10. It raises price and reduces quantity.

An important distinction between quotas and tariffs is that quotas do not increase costs to foreign producers; tariffs do. In the short run, a tariff will reduce the profits of foreign exporters of a good or service. A quota, however, raises price but not costs of production and thus may increase profits. Because the quota imposes a limit on quantity, any profits it creates in other countries will not induce the entry of new firms that ordinarily eliminates profits in perfect competition. By definition, entry of new foreign firms to earn the profits available in the United States is blocked by the quota.

3.4 Voluntary Export Restrictions

Voluntary export restrictions are a form of trade barrier by which foreign firms agree to limit the quantity of goods exported to a particular country. They became prominent in the United States in the 1980s, when the U.S. government persuaded foreign exporters of automobiles and steel to agree to limit their exports to the United States.

voluntary export restrictions

A form of trade barrier by which foreign firms agree to limit the quantity of goods exported to a particular country.

Although such restrictions are called voluntary, they typically are agreed to only after pressure is applied by the country whose industries they protect. The United States, for example, has succeeded in pressuring many other countries to accept quotas limiting their exports of goods ranging from sweaters to steel.

A voluntary export restriction works precisely like an ordinary quota. It raises prices for the domestic product and reduces the quantity consumed of the good or service affected by the quota. It can also increase the profits of the firms that agree to the quota because it raises the price they receive for their products.

3.5 Other Barriers

In addition to tariffs and quotas, measures such as safety standards, labeling requirements, pollution controls, and quality restrictions all may have the effect of restricting imports.

Many restrictions aimed at protecting consumers in the domestic market create barriers as a purely unintended, and probably desirable, side effect. For example, limitations on insecticide levels in foods are often more stringent in the United States than in other countries. These standards tend to discourage the import of foreign goods, but their primary purpose appears to be to protect consumers from harmful chemicals, not to restrict trade. But other nontariff barriers seem to serve no purpose other than to keep foreign goods out. Tomatoes produced in Mexico, for example, compete with those produced in the United States. But Mexican tomatoes tend to be smaller than U.S. tomatoes. The United States once imposed size restrictions to "protect" U.S. *consumers* from small tomatoes. The result was a highly effective trade barrier that protected U.S. producers and raised U.S. tomato prices. Those restrictions were abolished under terms of the North American Free Trade Agreement, which has led to a large increase in U.S. imports of Mexican tomatoes and a reduction in U.S. tomato production.[2]

3.6 Justifications for Trade Restriction: An Evaluation

The conceptual justification for free trade is one of the oldest arguments in economics; there is no disputing the logic of the argument that free trade increases global production, worldwide consumption, and international efficiency. But critics stress that the argument is a theoretical one. In the real world, they say, there are several arguments that can be made to justify protectionist measures.

Infant Industries

One argument for trade barriers is that they serve as a kind of buffer to protect fledgling domestic industries. Initially, firms in a new industry may be too small to achieve significant economies of scale and could be clobbered by established firms in other countries. A new domestic industry with potential economies of scale is called an **infant industry.**

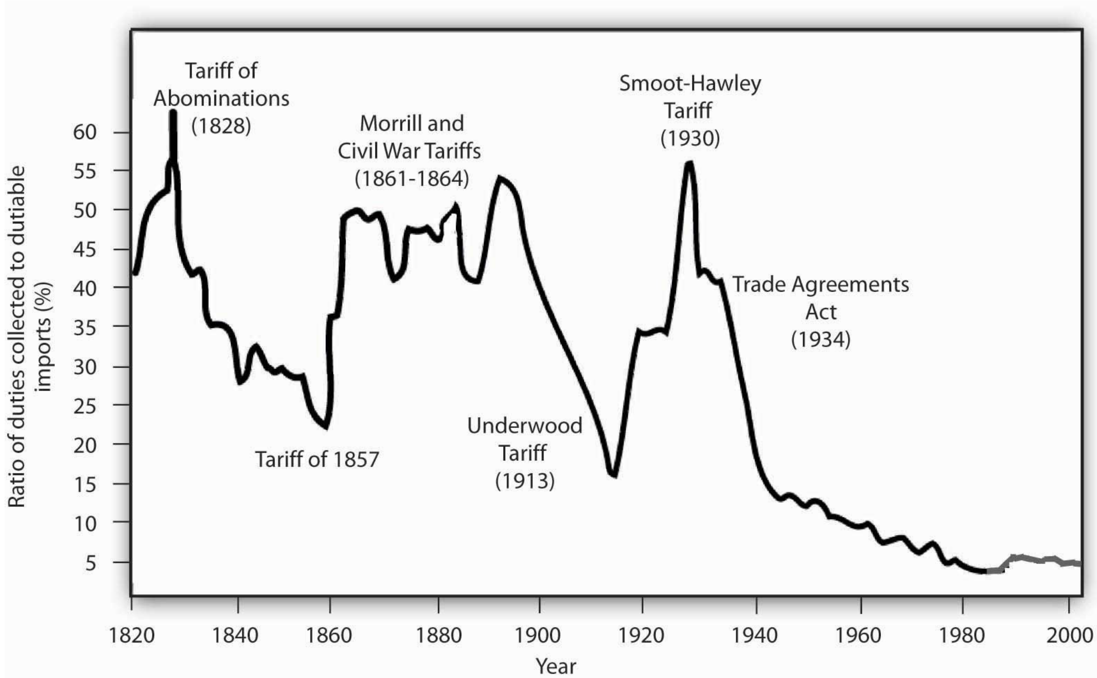
infant industry

A new domestic industry with potential economies of scale.

Consider the situation in which firms in a country are attempting to enter a new industry in which many large firms already exist in the international arena. The foreign firms have taken advantage of economies of scale and have therefore achieved relatively low levels of production costs. New firms, facing low levels of output and higher average costs, may find it difficult to compete. The infant industry argument suggests that by offering protection during an industry's formative years, a tariff or quota may allow the new industry to develop and prosper.

FIGURE 17.11 U.S. Tariff Rates, 1820–2005

Tariff rates on "dutiable imports" have fallen dramatically over the course of U.S. history.



Sources: Historical Statistics, Colonial Times to 1970; Statistical Abstract of the United States, 1998, Table no. 1325; Statistical Abstract of the United States, 1990; U.S. International Commission http://dataweb.usitc.gov/prepared_reports.asp).

The infant industry argument played a major role in tariff policy in the early years of U.S. development. Figure 17.11 shows average tariff rates on dutiable imports in the United States since 1820. The high tariffs of the early nineteenth century were typically justified as being necessary to allow U.S. firms to gain a competitive foothold in the world economy. As domestic industries became established, tariff rates fell. Subsequent increases in tariffs were a response in part to internal crises: the Civil War and the Great Depression. Tariff rates have fallen dramatically since 1930.

Critics of the infant industry argument say that once protection is in place, it may be very difficult to remove. Inefficient firms, they contend, may be able to survive for long periods under the umbrella of infant industry protection.

Strategic Trade Policy

A new version of the infant industry argument has been used in the past few years as technological developments have spawned whole new industries and transformed existing ones. The new version of the infant industry argument assumes an imperfectly competitive market.

Suppose technological change has given rise to a new industry. Given the economies of scale in this industry, only a few firms are likely to dominate it worldwide—it will likely emerge as an oligopoly. The firms that dominate the industry are likely to earn economic profits that will persist. Furthermore, because there will be only a few firms, they will be located in only a few countries. Their governments could conceivably impose taxes on these firms' profits that would enhance economic well-being within the country. The potential for such gains may justify government efforts to assist firms seeking to acquire a dominant position in the new industry.

strategic trade policy

A policy aimed at promoting the development of key industries that may increase a countrys domestic well-being through trade with the rest of the world.

Government aid could take the form of protectionist trade policies aimed at allowing these firms to expand in the face of foreign competition, assistance with research and development efforts, programs to provide workers with special skills needed by the industry, or subsidies in the form of direct payments or special tax treatment. Any such policy aimed at promoting the development of key industries that may increase a country's domestic well-being through trade with the rest of the world is known as a **strategic trade policy.**

Although strategic trade policy suggests a conceptually positive role for government in international trade, proponents of the approach note that it has dangers. Firms might use the strategic trade argument even if their development were unlikely to offer the gains specified in the theory. The successful application of the approach requires that the government correctly identify industries in which a country can, in fact, gain dominance—something that may not be possible. Various European governments provided subsidies to firms that were involved in the production of Airbus, which is now a major competitor in the airplane industry. On the other hand, Britain and France subsidized the development of the supersonic plane called the Concorde. After only a few Concordes had been produced, it became obvious that the aircraft was a financially losing proposition and production was halted. The airline has now gone out of business.

Finally, those firms whose success strategic trade policy promotes might have sufficient political clout to block the taxes that would redistribute the gains of the policies to the population in general. Thus, the promise of strategic trade policy is unlikely to be fulfilled.

National Security

It is sometimes argued that the security of the United States would be threatened if this country depended on foreign powers as the primary source of strategic materials. In time of war, the United States might be cut off from sources of foreign supply and lose some of the materials upon which U.S. industry depends.

One area where the national security argument is applied is the oil industry. Given the volatility of the political situation in the Middle East, some people say, the United States should protect the domestic oil industry in order to ensure adequate production capability in the event Middle Eastern supplies are cut off.

An alternative to tariff protection of strategic commodities is to stockpile those commodities for use in time of crisis. For example, the United States maintains a strategic petroleum reserve for use in case of a cutoff in foreign supplies or domestic crises. For example, strategic oil reserves were tapped in the wake of pipeline and refinery disruptions following Hurricane Katrina in 2005.

Job Protection

The desire to maintain existing jobs threatened by foreign competition is probably the single most important source of today's protectionist policies. Some industries that at one time had a comparative advantage are no longer among the world's lowest-cost producers; they struggle to stay afloat. Cost cutting leads to layoffs, and layoffs lead to demands for protection.

The model of international trade in perfect competition suggests that trade will threaten some industries. As countries specialize in activities in which they have a comparative advantage, sectors in which they do not have this advantage will shrink. Maintaining those sectors through trade barriers blocks a nation from enjoying the gains possible from free trade.

A further difficulty with the use of trade barriers to shore up employment in a particular sector is that it can be an enormously expensive strategy. Suppose enough of a foreign good is kept out of the United States to save one U.S. job. That shifts the supply curve slightly to the left, raising prices for U.S. consumers and reducing their consumer surplus. The loss to consumers is the cost per job saved. Estimates of the cost of saving *one* job in the steel industry through restrictions on steel imports, for example, go as high as $800,000 per year.

Cheap Foreign Labor and Outsourcing

One reason often given for the perceived need to protect American workers against free international trade is that workers must be protected against cheap foreign labor. This is an extension of the job protection argument in the previous section. From a theoretical point of view, of course, if foreign countries can produce a good at lower cost than we can, it is in our collective interest to obtain it from them. But workers counter by saying that the low wages of foreign workers means that foreign workers are exploited. To compete with foreign workers, American workers would have to submit themselves to similar exploitation. This objection, however, fails to recognize that differences in wage rates generally reflect differences in worker productivity.

Consider the following example: Suppose U.S. workers in the tool industry earn $20 per hour while Indonesian workers in the tool industry earn only $2 per hour. If we assume that the tool industry is competitive, then the wages in both countries are based on the marginal revenue product of the workers. The higher wage of U.S. workers must mean that they have a higher marginal product—they are more productive. The higher wage of U.S. workers need not mean that labor costs are higher in the United States than in Indonesia.

Further, we have seen that what matters for trade is comparative advantage, not comparative labor costs. When each nation specializes in goods and services in which it has a comparative advantage—measured in the amounts of other goods and services given up to produce them—then world production, and therefore world consumption, rises. By definition, each nation will have a comparative advantage in *something.*

outsourcing

Situation in which firms in a developed country transfer some of their activities abroad in order to take advantage of lower labor costs in other countries.

A particularly controversial issue in industrialized economies is **outsourcing**, in which firms in a developed country transfer some of their activities abroad in order to take advantage of lower labor costs in other countries. Generally speaking, the practice of outsourcing tends to reduce costs for the firms that do it. These firms often expand production and increase domestic employment, as is discussed in the accompanying Case in Point essay.

Differences in Environmental Standards

Another justification for protectionist measures is that free trade is unfair if it pits domestic firms against foreign rivals who do not have to adhere to the same regulatory standards. In the debate over NAFTA, for example, critics warned that Mexican firms, facing relatively lax pollution control standards, would have an unfair advantage over U.S. firms if restraints on trade between the two countries were removed.

Economic theory suggests, however, that differences in pollution-control policies can be an important source of comparative advantage. In general, the demand for environmental quality is positively related to income. People in higher-income countries demand higher environmental quality than do people in lower-income countries. That means that pollution has a lower cost in poorer than in richer countries. If an industry generates a great deal of pollution, it may be more efficient to locate it in a poor country than in a rich country. In effect, a poor country's lower demand for environmental quality gives it a comparative advantage in production of goods that generate a great deal of pollution.

Provided the benefits of pollution exceed the costs in the poor country, with the costs computed based on the preferences and incomes of people in that country, it makes sense for more of the good to be produced in the poor country and less in the rich country. Such an allocation leaves people in both countries better off than they would be otherwise. Then, as freer trade leads to higher incomes in the poorer countries, people there will also demand improvements in environmental quality.

Do economists support *any* restriction on free international trade? Nearly all economists would say no. The gains from trade are so large, and the cost of restraining it so high, that it is hard to find any satisfactory reason to limit trade.

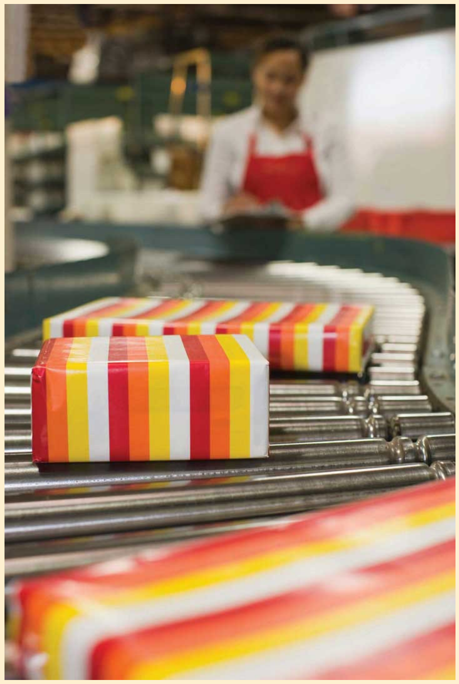
KEY TAKEAWAYS

- Protectionist measures seek to limit the quantities of goods and services imported from foreign countries. They shift the supply curve for each of the goods or services protected to the left.
- The primary means of protection are tariffs and quotas.
- Antidumping proceedings have emerged as a common means of protection.
- Voluntary export restrictions are another means of protection; they are rarely voluntary.
- Other protectionist measures can include safety standards, restrictions on environmental quality, labeling requirements, and quality standards.
- Protectionist measures are sometimes justified using the infant industry argument, strategic trade policy, job protection, "cheap" foreign labor and outsourcing, national security, and differences in environmental standards.

TRY IT!

Suppose the United States imposes a quota reducing its imports of shoes by one-half (roughly 85–90% of the shoes now sold in the United States are imported). Assume that shoes are produced under conditions of perfect competition and that the equilibrium price of shoes is now $50 per pair. Illustrate and explain how this quota will affect the price and output of shoes in the United States.

Case in Point: Outsourcing and Employment



The phenomenon of outsourcing has become common as the Internet and other innovations in communication have made it easier for firms to transfer aspects of their production overseas. At the same time, countries such as India and China have invested heavily in education and have produced a sizable workforce of professional people capable of filling relatively high level positions for firms in more developed countries.

The very idea of outsourcing rankles politicians on the left and on the right. In the United States, there have been numerous congressional hearings on outsourcing and proposals to block firms that engage in the practice from getting government contracts.

By outsourcing, firms are able to reduce their production costs. As we have seen, a reduction in production costs translates into increased output and falling prices. From a consumer's point of view, then, outsourcing should be a very good thing. The worry many commentators express, however, is that outsourcing will decimate employment in the United States, particularly among high-level professionals. Matthew J. Slaughter, an economist at Dartmouth University, examined employment trends from 1991 to 2001 among multinational U.S. firms that had outsourced jobs. Those firms outsourced 2.8 million jobs during the period.

Were the 2.8 million jobs simply lost? Mr. Slaughter points out that there are three reasons to expect that the firms that reduced production costs by outsourcing would actually increase their domestic employment. First, by lowering cost, firms are likely to expand the quantity they produce. The foreign workers who were hired, who Mr. Slaughter refers to as "affiliate workers," appeared to be complements to American workers rather than substitutes. If they are complements rather than substitutes, then outsourcing could lead to increased employment in the country that does the outsourcing.

A second reason outsourcing could increase employment is that by lowering production cost, firms that increase the *scale* of their operations through outsourcing need more domestic workers to sell the increased output, to coordinate its distribution, and to develop the infrastructure to handle all those goods.

Finally, firms that engage in outsourcing are also likely to increase the *scope* of their operations. They will need to hire additional people to explore other product development, to engage in research, and to seek out new markets for the firm's output.

Thus, Mr. Slaughter argues that outsourcing may lead to increased employment because domestic workers are complements to foreign workers, because outsourcing expands the scale of a firm's operations, and because it expands the scope of operations. What did the evidence show? Remember the 2.8 million jobs that multinational firms based in the United States outsourced between 1991 and 2001? Employment at those same U.S. firms *increased* by 5.5 million jobs during the period. Thus, with the phenomena of complementarity, increases in scale, and increases of scope, each job outsourced led to almost two *additional* jobs in the United States.

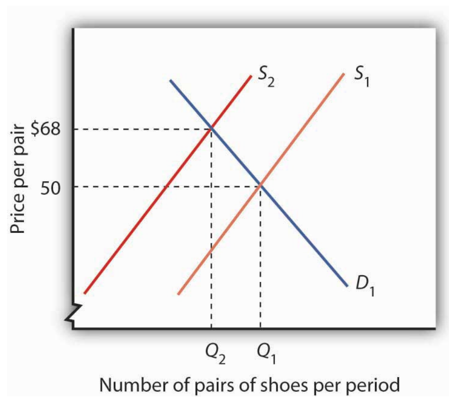
The experience of two quite dissimilar firms illustrates the phenomenon. Wal-Mart began expanding its operations internationally in about 1990. Today, it manages its global operations from its headquarters in Bentonville, Arkansas where it employs 15,000 people. Roughly 1,500 of these people coordinate the flow of goods among Wal-Mart's stores throughout the world. Those 1,500 jobs would not exist were it not for globalization. Xilinx, the high technology research and development firm, generates sales of about $1.5 billion per year. Sixty-five percent of its sales are generated outside the United States. But 80% of its employees are in the United States.

Outsourcing, then, generates jobs. It does not destroy them. Mr. Slaughter concludes: "Instead of lamenting ongoing foreign expansion of U.S. multinationals, if history is our guide then we should be encouraging it."

Source: Matthew J. Slaughter, "Globalization and Employment by U.S. Multinationals: A Framework and Facts," Daily Tax Report (March 26, 2004): 1–12.

ANSWER TO TRY IT! PROBLEM

The quota shifts the supply curve to the left, increasing the price of shoes in the United States and reducing the equilibrium quantity. In the case shown, the price rises to $68. Because you are not given the precise positions of the demand and supply curves, you can only conclude that price rises; your graph may suggest a different price. The important thing is that the new price is greater than $50.



4. REVIEW AND PRACTICE

Summary

In this chapter we have seen how international trade makes it possible for countries to improve on their domestic production possibilities.

A country that is operating on its production possibilities curve can obtain more of all goods by opening its markets to free international trade. Free trade allows nations to consume goods beyond their domestic production possibilities curves. If nations specialize in the production of goods and services in which they have a comparative advantage, total output increases. Free trade enhances production possibilities on a worldwide scale. It does not benefit everyone, however. Some workers and owners of other factors of production will be hurt by free trade, at least in the short run.

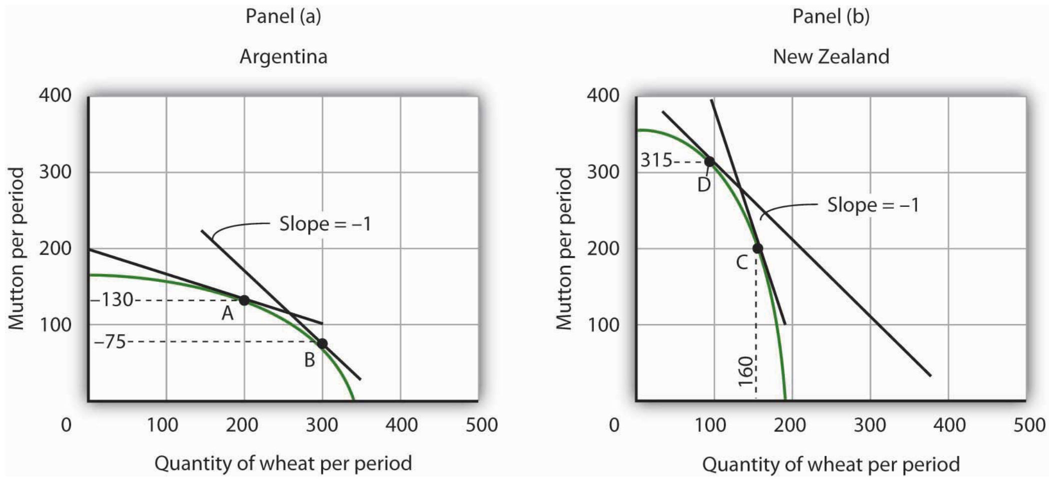
Contrary to the implication of the model of specialization based on comparative advantage, not all trade is one-way trade. Two-way trade in the *same* goods may arise from variations in transportation costs and seasonal influences. Two-way trade in *similar* goods is often the result of imperfect competition. Much trade among high-income countries is two-way trade.

The imposition of trade barriers such as tariffs, antidumping proceedings, quotas, or voluntary export restrictions raises the equilibrium price and reduces the equilibrium quantity of the restricted good. Although there are many arguments in favor of such restrictions on free trade, economists generally are against protectionist measures and supportive of free trade.

CONCEPT PROBLEMS

1. Explain how through trade a country can consume at levels beyond the reach of its production possibilities.
2. Why do countries place restrictions on international trade?
3. What is the difference between a tariff and a quota?
4. The Case in Point on America's shifting comparative advantage suggests that the United States may have a comparative advantage over other countries in the production of high-tech capital goods. What do you think might be the sources of this advantage?
5. "I know a lawyer who can type 100 words per minute but pays a secretary $10 per hour to type court briefs. But the secretary can only type 50 words per minute. I have told my lawyer friend a hundred times she would be better off doing the typing herself, but she just will not listen." Who has the better part of this disagreement, the lawyer or the friend? Explain.
6. Which individuals in the United States might benefit from a tariff placed on the importation of shoes? Who might lose?
7. Explain why economists argue that protectionist policies lead to the misallocation of resources in the domestic economy.
8. Tomatoes grow well in Kansas. Why do the people of Kansas buy most of their tomatoes from Florida, Mexico, and California?
9. Under what circumstances will a country both export and import the products of the same industry?
10. Suppose the United States imposes a quota on copper imports. Who might be helped? Who might be hurt?
11. Some people argue that international trade is fine, but that firms in different countries should play on a "level playing field." They argue that if a good can be produced more cheaply abroad than at home, tariffs should be imposed on the good so that the costs of producing it are the same everywhere. What do you think of this argument?
12. Suppose wages in the Philippines are one-tenth of wages in the United States. Why do all U.S. firms not just move production to the Philippines?

NUMERICAL PROBLEMS



1. Argentina and New Zealand each produce wheat and mutton under conditions of perfect competition, as shown on the accompanying production possibilities curves. Assume that there is no trade between the two countries and that Argentina is now producing at point A and New Zealand at point C.
    a. What is the opportunity cost of producing each good in Argentina?
    b. What is the opportunity cost of producing each good in New Zealand?
    c. Which country has a comparative advantage in which good? Explain.
    d. Explain how international trade would affect wheat production in Argentina.
    e. How would international trade affect mutton production?
    f. Explain how international trade would affect wheat production in New Zealand. How would it affect mutton production?
    g. How would trade between the two countries affect consumption of wheat and mutton in each country?
2. Assume that trade opens between Argentina and New Zealand and that, with trade, a pound of mutton exchanges for a bushel of wheat. Before trade, Argentina produced at point A and New Zealand produced at point C. Argentina moves to point B, while New Zealand moves to point D. Calculate and illustrate graphically an exchange between Argentina and New Zealand that would leave both countries with more of both goods than they had before trade.
3. Assume that the world market for producing radios is monopolistically competitive. Suppose that the price of a typical radio is $25.
    a. Why is this market likely to be characterized by two-way trade?
    b. Suppose that Country A levies a tax of $5 on each radio produced within its borders. Will radios continue to be produced in Country A? If they are, what will happen to their price? If they are not, who will produce them?
    c. If you concluded that radios will continue to be produced in Country A, explain what will happen to their price in the short run. Illustrate your answer graphically.
    d. What will happen to their price in the long run?
4. Suppose radio producers in Country A file a successful anti-dumping complaint against their competitors, and that the result is the imposition of a $10 per radio tariff on imported radios.
    a. Illustrate and explain how the $10 tariff will affect radio producers in Country A in the *short* run.
    b. Illustrate and explain how the $10 tariff will affect radio producers in Country A in the *long* run.
    c. How will the level of employment be affected in Country A?
    d. Explain how the tariff will affect consumers in Country A. Who will benefit from the anti-dumping action? Who will bear the burden of the action?

ENDNOTES

1. "Who Wants to Be a Millionaire?" *The Wall Street Journal*, May 14, 2008, p. A20.

2. Ramon G. Guajardo and Homero A. Elizondo, "North American Tomato Market: A Spatial Equilibrium Perspective," *Applied Economics*, 35(3) (February 2003): 315–22.

CHAPTER 18

The Economics of the Environment

START UP: A MARKET FOR CARBON EMISSIONS

Emissions of carbon dioxide, a "greenhouse gas," are limited by terms of the Kyoto accords, a treaty signed and ratified by most of the world's nations. The treaty requires industrialized nations ratifying it to reduce their emissions of greenhouse gases such as carbon dioxide, which are thought to cause global warming, by 5.2% below their 1990 levels. The treaty also requires industrialized nations to assist developing nations with their efforts to reduce emissions of greenhouse gases. The vast majority of the world's nations have ratified the accords. The United States is among a minority of nations that are not participating in them.

The Kyoto accords require that each industrialized nation specify quotas for each firm or organization that emits greenhouse gases. Those that emit less than their quotas can sell their unused rights to emit gases to other firms or organizations. Any agent that wishes to exceed its quota must purchase rights from someone else. That exchange permits organizations that can meet their quotas at relatively low cost to sell rights to others for which reductions would be costly. The exchanges still allow the overall target to be met, but at much lower cost.

President George W. Bush's predecessor, President Bill Clinton, signed the Kyoto agreement, but it was not ratified by the United States Senate. President Bush withdrew the United States from the agreement citing the damage to the economy that implementation would cause, the fact that rapidly industrializing economies, such as China and India, are excluded, and inconclusive science. Despite this, many firms and other organizations in the United States have already agreed to limit their emissions of greenhouse gases. For the more than 100 members of the Chicago Climate Exchange (CCX), companies and organizations including IBM, the Ford Motor Company, Motorola, and the cities of Chicago and Oakland, the idea of exchanging rights to dump greenhouse gases in the atmosphere is already a reality.

Members of the CCX, formed in 2003 by Richard Sandor, a commodities entrepreneur and economist who helped establish markets in financial futures, have pledged to reduce their emissions of greenhouse gases to an annual target. If a member wishes to undertake a project that would generate additional carbon emissions, it purchases rights to the additional carbon from another member. In its first nine months of operation, CCX members traded rights to more than 1,000,000 tons of carbon. The CCX quotas are far less stringent than are those of the Kyoto accords. The chief importance of CCX is that it establishes a model for the exchange of emissions rights should the United States sign the treaty.

The development of a market for pollution rights puts market forces to work in an effort to reduce the potential problem of global warming. The European Union introduced exchanges in rights to emit greenhouse gases and issued quotas to firms for carbon dioxide emissions in its 25 member nations in 2005. Anticipating the limits, firms began trading rights to emit carbon dioxide in 2004, when the European Climate Exchange was announced. The price of the right to dump a metric ton of greenhouse gases was about $20. By 2008, the price had risen to €29, and rights to emit 295 million tons were traded on the ECX in June of that year.[1] In the purely voluntary Chicago Climate Exchange, rights were traded much more cheaply—for about $5 per metric ton.[2] The CCX has since formed a partnership with London's International Petroleum Exchange and the European Climate Exchange, so that European firms can exchange carbon rights on an international basis.

In this chapter we shall put the analytical tools we have developed to work on the problems of the environment. We will begin with an examination of problems of air and water pollution. We will look also at alternative regulatory approaches to environmental problems. Direct controls, in which government agencies tell polluters what they must do to reduce pollution, remain the primary means of government regulation. Persuasion, seeking voluntary compliance with pollution reduction efforts, is also used. Two alternatives that economists advocate are taxes on pollutants and marketable pollution rights; such systems are gaining in importance.

Related issues include those of common property resources, those scarce resources for which no property rights are defined, and exhaustible natural resources, those resources whose stocks decline as they are used. These topics were discussed in the chapters on efficiency and natural resources, respectively.

1. MAXIMIZING THE NET BENEFITS OF POLLUTION

LEARNING OBJECTIVES

1. **Explain why pollution can be said to have benefits as well as costs and describe the nature of these benefits and costs.**
2. **Using marginal benefit and marginal cost curves, apply the marginal decision rule to show and explain what is meant by the efficient level of emissions and abatement.**
3. **Explain the Coase theorem and what it implies about the conditions under which the private market is likely to achieve an efficient outcome.**

We all pollute the environment. We do so not because we get some perverse satisfaction from polluting, but because activities that give us utility inevitably pollute. We do not drive our cars in order to dump carbon monoxide into the air but because we gain utility from the transportation and convenience cars provide. Firms pollute the environment if doing so allows them to produce goods and services at lower cost.

The benefits we derive from pollution are indirect. We obtain them from other activities that generate pollution. But that is not unusual—there are many things we do because of the other benefits they produce. Firms benefit from hiring labor not because their owners enjoy hiring workers but because those workers produce greater profits. We purchase electricity not because we enjoy the feeling of having the stuff racing through wires in the house but because the electricity produces light, heat, and other services more cheaply than would alternatives such as candles or fires. In purchasing this electricity, of course, we are demanding a good whose production inevitably degrades the environment. We pollute in the process of obtaining more of other goods and services we enjoy. We thus benefit from our pollution.

Of course, we suffer from the pollution we all generate as well. Smog-choked air damages our health and robs us of scenic views. We may not be able to fish or swim in polluted rivers. Just as the generation of pollution makes many of the activities we pursue less expensive, the fact that we have pollution increases many costs. Polluted rivers increase the cost of producing drinking water. Polluted air requires us to spend more on health care and to paint our buildings more often. Polluted soils produce less food.

Like any other activity, then, pollution has benefits as well as costs. The difficulty with pollution problems is that decision makers experience the benefits of their own choices to pollute the environment, but the costs spill over to everyone who breathes the air or consumes the water. These costs are examples of external costs. Recall that external costs produce one type of market failure and that market failures lead to inefficiency in the allocation of resources. The environment presents us with an allocation problem in which decision makers are not faced with all the benefits and costs of their choices. Environmental resources will not, in such cases, be allocated efficiently. Economists who examine and analyze environmental problems try to determine what an efficient allocation of the environment would most likely be—one that maximizes the difference between the total benefits and total costs of our pollution.

A second task of environmental economics is to find ways to get from where we are, typically with more pollution than is efficient, to the efficient solution. We have learned that private markets often fail to achieve efficient solutions to environmental problems because property rights are difficult to define and to exchange. We will see, however, that environmental economists have devised innovative ways to introduce property rights to environmental policy and to harness market forces to improve rather than degrade environmental quality.

1.1 Pollution and Scarcity

Pollution exists whenever human activity generates a sufficient concentration of a substance in the environment to cause harm to people or to resources valued by people. Many potentially harmful substances are natural features of the environment, but they are not generally regarded as pollutants. Pollution is the product of people, not nature.

pollution

Situation that exists when human activity produces a sufficient concentration of a substance in the environment to cause harm to people or to resources valued by people.

Pollution implies scarcity. If an activity emits harmful by-products into the environment, then the emission of the by-products is an alternative to some other activity. A scarcity problem exists at the point where harm occurs. A fire burning in a fireplace at a cabin in the forest whose smoke goes unnoticed does not suggest a scarcity problem. But when there are other people who will be harmed by the smoke, then one person's enjoyment of the fire becomes an alternative to another person's enjoyment of fresh air. Fresh air has become scarce, and pollution has become an economic problem.

Economists generally argue that pollution that harms plants or animals imposes a cost if the plants or animals are valued by people. When a farmer uses a pesticide that damages another farmer's crop, for example, a pollution problem occurs. If an oil spill in the ocean damages sea animals that people care about, there is a pollution problem. It is, after all, people who make the choices that lead to pollution. It is people who can choose to limit their pollution. Economists therefore examine pollution problems from the perspective of the preferences of people.

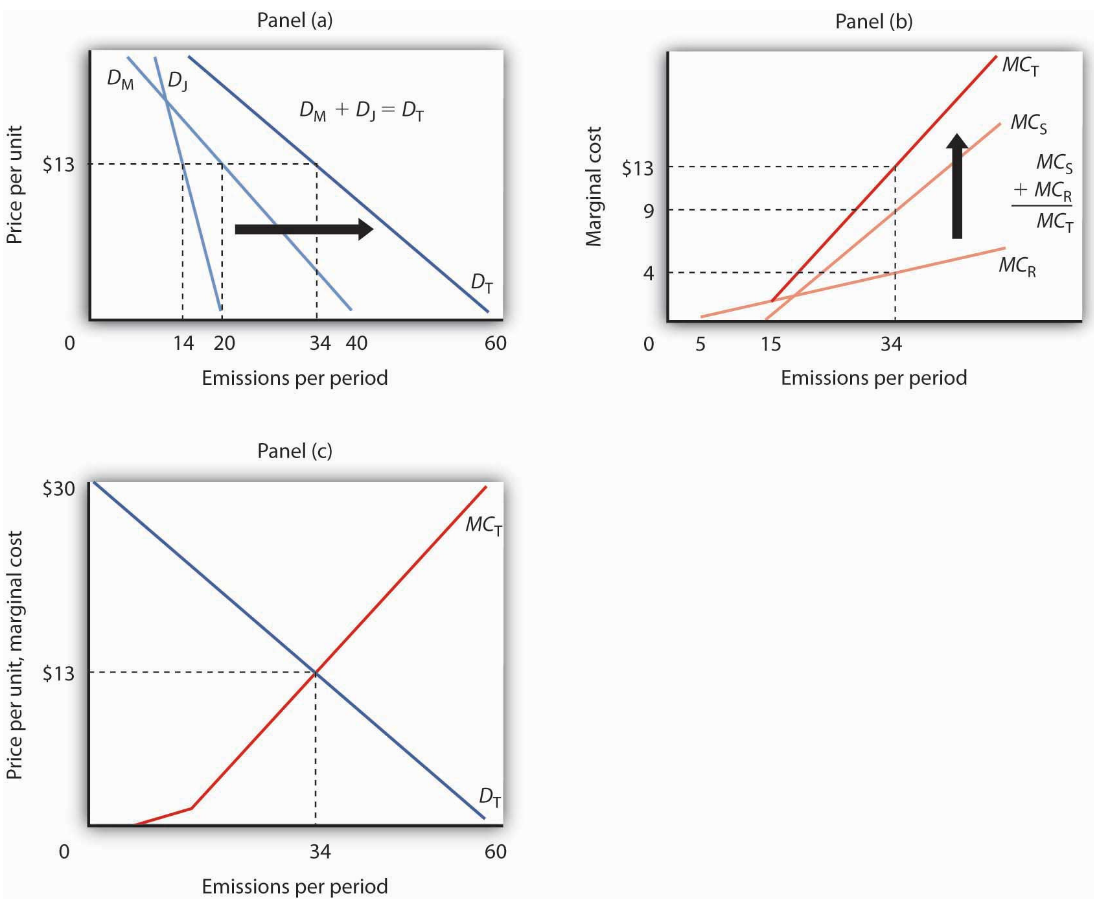
1.2 The Efficient Level of Pollution

The efficient level of pollution is the quantity at which its total benefits exceed its total costs by the greatest possible amount. This occurs where the marginal benefit of an additional unit of pollution equals its marginal cost.

Figure 18.1 shows how we can determine an efficient quantity of pollution. Suppose two neighbors in a remote mountain area, Mary and Jane, burn fires in their cabins that generate air pollution that harms two other individuals, Sam and Richard, who live downwind. We shall assume that Mary and Jane are the only polluters and that Sam and Richard are the only people harmed by the pollution. We shall ignore, for example, the effect of Mary and Jane's emissions on the possible problem of global warming.

FIGURE 18.1 Determining the Efficient Level of Pollution

Mary and Jane each benefit from polluting the environment by emitting smoke from their fires. In Panel (a), we see that Mary's demand curve for emitting smoke is given by D_M and that Jane's demand curve is given by D_J. To determine the total demand curve, D_T, we determine the amount that each person will emit at various prices. At a price of $13 per day, for example, Mary will emit 20 pounds per day. Jane will emit 14 pounds per day, for a total of 34 pounds per day on demand curve D_T. Notice that if the price were $0 per unit, Mary would emit 40 pounds per day, Jane would emit 20, and emissions would total 60 pounds per day on curve D_T. The marginal cost curve for pollution is determined in Panel (b) by taking the marginal cost curves for each person affected by the pollution. In this case, the only people affected are Sam and Richard. Sam's marginal cost curve is MC_S, and Richard's marginal cost curve is MC_R. Because Sam and Richard are each affected by the same pollution, we add their marginal cost curves vertically. For example, if the total quantity of the emissions is 34 pounds per day, the marginal cost of the 34th pound to Richard is $4, it is $9 to Sam, for a total marginal cost of the 34th unit of $13. In Panel (c), we put D_T and MC_T together to find the efficient solution. The two curves intersect at a level of emissions of 34 pounds per day, which occurs at a price of $13 per pound.



Suppose, as is often the case, that no mechanism exists to charge Mary and Jane for their emissions; they can pollute all they want and never have to compensate society (that is, pay Sam and Richard) for the damage they do. Alternatively, suppose there is no mechanism for Sam and Richard to pay Mary and Jane to get them to reduce their pollution. In either situation, the pollution generated by Mary and Jane imposes an external cost on Sam and Richard. Mary and Jane will pollute up to the point that the marginal benefit of additional pollution to them has reached zero—that is, up to the point where the marginal benefit matches their marginal cost. They ignore the external costs they impose on "society"—Sam and Richard.

Mary's and Jane's demand curves for pollution are shown in Panel (a). These demand curves, D_M and D_J, show the quantities of emissions each generates at each possible price, assuming such a fee were assessed. At a price of $13 per unit, for example, Mary will emit 20 units of pollutant per period and Jane will emit 14. Total emissions at a price of $13 would be 34 units per period. If the price of emissions were zero, total emissions would be 60 units per period. Whatever the price they face, Mary and Jane will emit additional units of the pollutant up to the point that their marginal benefit equals that price. We can therefore interpret their demand curves as their marginal benefit curves for emissions. Their combined demand curve D_T gives the marginal benefit to society (that is, to Mary and

Jane) of pollution. Each person in our problem, Mary, Jane, Sam, and Richard, follows the marginal decision rule and thus attempts to maximize utility.

In Panel (b) we see how much Sam and Richard are harmed; the marginal cost curves, MC_S and MC_R, show their respective valuations of the harm imposed on them by each additional unit of emissions. Notice that over a limited range, some emissions generate no harm. At very low levels, neither Sam nor Richard is even aware of the emissions. Richard begins to experience harm as the quantity of emissions goes above 5 pounds per day; it is here that pollution begins to occur. As emissions increase, the additional harm each unit creates becomes larger and larger—the marginal cost curves are upward sloping. The first traces of pollution may be only a minor inconvenience, but as pollution goes up, the problems it creates become more serious—and its marginal cost rises.

Because the same emissions affect both Sam and Richard, we add their marginal cost curves vertically to obtain their combined marginal cost curve MC_T. The 34th unit of emissions, for example, imposes an additional cost of $9 on Sam and $4 on Richard. It thus imposes a total marginal cost of $13.

The efficient quantity of emissions is found at the intersection of the demand (D_T) and marginal cost (MC_T) curves in Panel (c) of Figure 18.1, with 34 units of the pollutant emitted. The marginal benefit of the 34th unit of emissions, as measured by the demand curve D_T, equals its marginal cost, MC_T, at that level. The quantity at which the marginal benefit curve intersects the marginal cost curve maximizes the net benefit of an activity.

We have already seen that in the absence of a mechanism to charge Mary and Jane for their emissions, they face a price of zero and would emit 60 units of pollutant per period. But that level of pollution is inefficient. Indeed, as long as the marginal cost of an additional unit of pollution exceeds its marginal benefit, as measured by the demand curve, there is too much pollution; the net benefit of emissions would be greater with a lower level of the activity.

Just as too much pollution is inefficient, so is too little. Suppose Mary and Jane are not allowed to pollute; emissions equal zero. We see in Panel (c) that the marginal benefit of dumping the first unit of pollution is quite high; the marginal cost it imposes on Sam and Richard is zero. Because the marginal benefit of additional pollution exceeds its marginal cost, the net benefit to society would be increased by increasing the level of pollution. That is true at any level of pollution below 34 units, the efficient solution.

The notion that too little pollution could be inefficient may strike you as strange. To see the logic of this idea, imagine that the pollutant involved is carbon monoxide, a pollutant emitted whenever combustion occurs, and it is deadly. It is, for example, emitted when you drive a car. Now suppose that no emissions of carbon monoxide are allowed. Among other things, this would require a ban on all driving. Surely the benefits of some driving would exceed the cost of the pollution created. The problem in pollution policy from an economic perspective is to find the quantity of pollution at which total benefits exceed total costs by the greatest possible amount—the solution at which marginal benefit equals marginal cost.

1.3 Property Rights and the Coase Theorem

The problem of getting the efficient amount of pollution arises because no one owns the right to the air. If someone did, then that owner could decide how to use it. If a nonowner did not like the way the owner was using it, then he or she could try to pay the owner to change the way the air was being used.

In our earlier example, if Mary and Jane own the right to the air, but Sam and Richard do not like how much Mary and Jane are polluting it, Sam and Richard can offer to pay Mary and Jane to cut back on the amount they are polluting. Alternatively, if Sam and Richard own the air, then in order to pollute it, Mary and Jane would have to compensate the owners. Bargaining among the affected parties, if costless, would lead to the efficient amount of pollution. Costless bargaining requires that all parties know the source of the pollution and are able to measure the quantity emitted by each agent.

Specifically, suppose Mary and Jane own the right to dump pollutants into the air and had been emitting 60 units of the pollution per day. If Sam and Richard offer to pay them $13 for each unit of pollutant they cut back, Mary and Jane will reduce their pollution to 34 units, since the marginal benefit to them of the 35th to 60th unit is less than $13. Sam and Richard are better off since the marginal cost of those last 26 units of pollution is greater than $13 per unit. The efficient outcome of 34 units of pollution is thus achieved. Mary and Jane could reduce their emissions by burning their fires for less time, by selecting more efficient fireplaces, or by other measures. Many people, for example, have reduced their emissions by changing to gas-burning fireplaces rather than burning wood.

Coase theorem

The proposition that if property rights are well defined and if bargaining is costless, then the private market can achieve an efficient outcome regardless of which of the affected parties holds the property rights.

While the well-being of the affected parties is not independent of who owns the property right (each would be better off owning rather than not owning the right), the establishment of who owns the air leads to a solution that solves the externality problem and leads to an efficient market outcome. The proposition that if property rights are well defined and if bargaining is costless, the private market can achieve an efficient outcome regardless of which of the affected parties holds the property rights is known as the **Coase theorem**, named for the Nobel-Prize-winning economist Ronald Coase who generally is credited with this idea.

Suppose instead that Sam and Richard own the right to the air. They could charge Mary and Jane $13 per unit of pollution emitted. Mary and Jane would willingly pay for the right to emit 34 units of pollution, but not for any more, since beyond that amount the marginal benefit of each unit of pollution emitted is less than $13. Sam and Richard would willingly accept payments for 34 units of pollution at $13 each since the marginal cost to them for each of those units is less than $13.

In most cases, however, Coase stressed that the conditions for private parties to achieve an efficient outcome on their own are not present. Property rights may not be defined. Even if one party owns a right, enforcement may be difficult. Suppose Sam and Richard own the right to clean air but many people, not just two as in our example, contribute to polluting it. Would each producer that pollutes have to strike a deal with Sam and Richard? Could the government enforce all those deals? And what if there are many owners as well? Enforcement becomes increasingly difficult and striking deals becomes more and more costly. Finally, monitoring is extremely difficult in most environmental problems. While it is generally possible to detect the presence of pollutants, determining their source is virtually impossible. And determining who is harmed and by how much is a monumental undertaking in most pollution problems.

Nonetheless, it is the insight that the Coase theorem provides that has led economists to consider solutions to environmental problems that attempt to use the establishment of property rights and market mechanisms in various ways to bring about the efficient market outcome. Before considering alternative ways of controlling pollution, we look first at how the benefits and costs might be measured so that we have a better sense of what the efficient solution is.

Another insight that comes from Coase's analysis is that the notion of "harm" is a reciprocal one. In our first example, it is tempting to conclude that Mary and Jane, by burning fires in their fireplaces, are "harming" Sam and Richard. But if Sam and Richard were not located downwind of the smoke, there would be no harm. In effect, Mr. Coase insists that the harm cannot be attributed to one party or another. Sam and Richard "cause" the harm by locating downwind of the fireplaces. While there is clearly harm in this situation, we could as easily attribute it to either the generators of the smoke or the recipients of the smoke. Before Coase wrote his article, "The Problem of Social Cost" in 1960, the general presumption was that decision makers such as Mary and Jane "cause" the harm and that they should be taxed for the costs they impose. Mr. Coase pointed out the alternative that Sam and Richard could avoid the harm by moving. Indeed, all sorts of alternative solutions come to mind even in this simple example. Mary and Jane could select types of wood that emit less smoke. They could use fireplaces that emit less smoke. They could make arrangements to time their burning to minimize the total amount of smoke that affects Sam and Richard. The goal, Coase said, is to select the most efficient from the alternatives available.[3]

Consider a different problem. Suppose that an airport has been built several miles outside the developed portion of a small city. No one lives close enough to the airport to be "harmed" by the noise inevitably generated by the operation of the airport. As time passes, people are likely to build houses near the airport to take advantage of jobs at the airport or to gain easy access to it. Those people will now be "harmed" by noise from the airport. But what is the cause of this harm? By definition, there was no "harm" before people started living close to the airport. True, it is the airport that generates the noise. But the noise causes harm *only* because people now live near it. Just as Mary and Jane's campfires only generate "harm" if someone is downwind of the smoke, the noise from the airport causes damage only if someone lives near the airport. The problem of the noise could be mitigated in several ways. First, people could have chosen not to live near the airport. Once they have chosen to live near the airport, they could reduce the noise with better insulation or with better windows. Alternatively, the airport management could choose different flight patterns to reduce noise that affects neighboring homeowners. It is always the case that there are several potential ways of mitigating the effects of airport operations; the economic problem is to select the most efficient from among those alternatives.

1.4 The Measurement of Benefits and Costs

Saying that the efficient level of pollution occurs at a certain rate of emissions, as we have done so far, is one thing. Determining the actual positions of the demand and marginal cost curves that define that efficient solution is quite another. Economists have devised a variety of methods for measuring these curves.

Benefits: The Demand for Emissions

A demand curve for emitting pollutants shows the quantity of emissions demanded per unit of time at each price. It can, as we have seen, be taken as a marginal benefit curve for emitting pollutants.

The general approach to estimating demand curves involves observing quantities demanded at various prices, together with the values of other determinants of demand. In most pollution problems, however, the price charged for emitting pollutants has always been zero—we simply do not know how the quantity of emissions demanded will vary with price.

One approach to estimating the demand curve for pollution utilizes the fact that this demand occurs because pollution makes other activities cheaper. If we know how much the emission of one more unit of a pollutant saves, then we can infer how much consumers or firms would pay to dump it.

Suppose, for example, that there is no program to control automobile emissions—motorists face a price of zero for each unit of pollution their cars emit. Suppose that a particular motorist's car emits an average of 10 pounds of carbon monoxide per week. Its owner could reduce emissions to 9 pounds per week at a cost of $1 per week. This $1 is the marginal cost of reducing emissions from 10 to 9 pounds per week. It is also the maximum price the motorist would pay to increase emissions from 9 to 10 pounds per week—it is the marginal benefit of the 10th pound of pollution. We say that it is the maximum price because if asked to pay more, the motorist would choose to reduce emissions at a cost of $1 instead.

Now suppose that emissions have been reduced to 9 pounds per week and that the motorist could reduce them to 8 at an additional cost of $2 per week. The marginal cost of reducing emissions from 9 to 8 pounds per week is $2. Alternatively, this is the maximum price the motorist would be willing to pay to increase emissions to 9 from 8 pounds; it is the marginal benefit of the 9th pound of pollution. Again, if asked to pay more than $2, the motorist would choose to reduce emissions to 8 pounds per week instead.

We can thus think of the marginal benefit of an additional unit of pollution as the added cost of not emitting it. It is the saving a polluter enjoys by dumping additional pollution rather than paying the cost of preventing its emission. Figure 18.2 shows this dual interpretation of cost and benefit. Initially, our motorist emits 10 pounds of carbon monoxide per week. Reading from right to left, the curve measures the marginal costs of pollution abatement (MC_A). We see that the marginal cost of abatement rises as emissions are reduced. That makes sense; the first reductions in emissions will be achieved through relatively simple measures such as modifying one's driving technique to minimize emissions (such as accelerating more slowly), or getting tune-ups more often. Further reductions, however, might require burning more expensive fuels or installing more expensive pollution-control equipment.

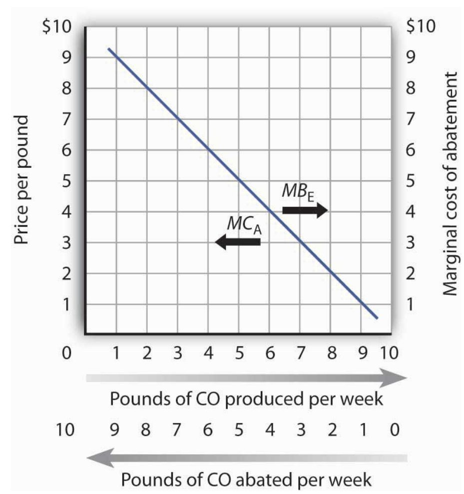
Read from left to right, the curve in Figure 18.2 shows the marginal benefit of additional emissions (MB_E). Its negative slope suggests that the first units of pollution emitted have very high marginal benefits, because the cost of not emitting them would be very high. As more of a pollutant is emitted, however, its marginal benefit falls—the cost of preventing these units of pollution becomes quite low.

Economists have also measured demand curves for emissions by using surveys in which polluters are asked to report the costs to them of reducing their emissions. In cases in which polluters are charged for the emissions they create, the marginal benefit curve can be observed directly.

As we saw in Figure 18.1, the marginal benefit curves of individual polluters are added horizontally to obtain a market demand curve for pollution. This curve measures the additional benefit to society of each additional unit of pollution.

FIGURE 18.2 Abatement Costs and Demand

A car emits an average of 10 pounds of CO per week when no restrictions are imposed—when the price of emissions is zero. The marginal cost of abatement (MC_A) is the cost of eliminating a unit of emissions; this is the interpretation of the curve when read from right to left. The same curve can be read from left to right as the marginal benefit of emissions (MB_E).



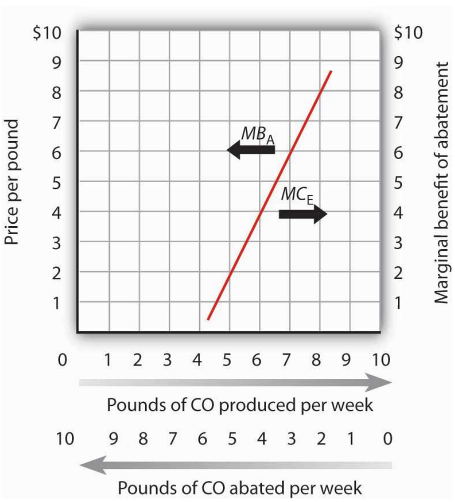
The Marginal Cost of Emissions

Pollutants harm people and the resources they value. The marginal cost curve for a pollutant shows the additional cost imposed by each unit of the pollutant. As we saw in Figure 18.1, the marginal cost curves for all the individuals harmed by a particular pollutant are added vertically to obtain the marginal cost curve for the pollutant.

Like the marginal benefit curve for emissions, the marginal cost curve can be interpreted in two ways, as suggested in Figure 18.3. When read from left to right, the curve measures the marginal cost of additional units of emissions (MC_E). If increasing the motorists' emissions from four pounds of carbon monoxide per week to five pounds of carbon monoxide per week imposes an external cost of $2, though, the marginal benefit of not being exposed to that unit of pollutant must be $2. The marginal cost curve can thus be read from right to left as a marginal benefit curve for abating emissions (MB_A). This marginal benefit curve is, in effect, the demand curve for cleaner air.

FIGURE 18.3 **The Marginal Cost of Emissions and the Marginal Benefit of Abatement**

The marginal cost of the first few units of emissions is zero and then rises once emissions begin to harm people. That is the point at which the air becomes a scarce resource. Read from left to right the curve gives the marginal cost of emissions (MC_E). Read from right to left, the curve gives the marginal benefit of abatement (MB_A).



Economists estimate the marginal cost curve of pollution in several ways. One is to infer it from the demand for goods for which environmental quality is a complement. Another is to survey people, asking them what pollution costs—or what they would pay to reduce it. Still another is to determine the costs of damages created by pollution directly.

For example, environmental quality is a complement of housing. The demand for houses in areas with cleaner air is greater than the demand for houses in areas that are more polluted. By observing the relationship between house prices and air quality, economists can learn the value people place on cleaner air—and thus the cost of dirtier air. Studies have been conducted in cities all over the world to determine the relationship between air quality and house prices so that a measure of the demand for cleaner air can be made. They show that increased pollution levels result in lower house values.[4]

Surveys are also used to assess the marginal cost of emissions. The fact that the marginal cost of an additional unit of emissions is the marginal benefit of avoiding the emissions suggests that surveys can be designed in two ways. Respondents can be asked how much they would be harmed by an increase in emissions, or they can be asked what they would pay for a reduction in emissions. Economists often use both kinds of questions in surveys designed to determine marginal costs.

A third kind of cost estimate is based on objects damaged by pollution. Increases in pollution, for example, require buildings to be painted more often; the increased cost of painting is one measure of the cost of added pollution.

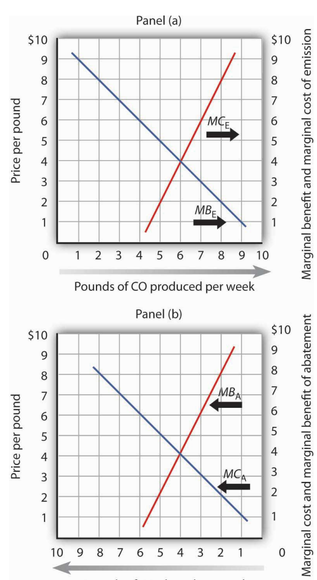
While all of the attempts to measure cost are imperfect, the alternative is to not try to quantify cost at all. To economists, such an ostrich-like approach of sticking one's head in the sand would be unacceptable.

The Efficient Level of Emissions and Abatement

Whether economists measure the marginal benefits and marginal costs of emissions or, alternatively, the marginal benefits and marginal costs of abatement, the policy implications are the same from an economic perspective. As shown in Panel (a) of Figure 18.4, applying the marginal decision rule in the case of emissions suggests that the efficient level of pollution occurs at six pounds of CO emitted per week. At any lower level, the marginal benefits of the pollution would outweigh the marginal costs. At a higher level, the marginal costs of the pollution would outweigh the marginal benefits.

FIGURE 18.4 **The Efficient Level of Emissions and Pollution Abatement**

In Panel (a) we combine the marginal benefit of emissions (MB_E) with the marginal cost of emissions (MC_E). The efficient solution occurs at the intersection of the two curves. Here, the efficient quantity of emissions is six pounds of CO per week. In Panel (b), we have the same curves read from right to left. The marginal cost curve for emissions becomes the marginal benefit of abatement (MB_A). The marginal benefit curve for emissions becomes the marginal cost of abatement (MC_A). With no abatement program, emissions total ten pounds of CO per week. The efficient degree of abatement is to reduce emissions by four pounds of CO per week to six pounds per week.



As shown in Panel (b) of Figure 18.4, application of the marginal decision rule suggests that the efficient level of *abatement* effort is to reduce pollution by 4 pounds of CO produced per week. That is, reduce the level of pollution from the 10 pounds per week that would occur at a zero price to 6 pounds per week. For any greater effort at abating the pollution, the marginal cost of the abatement efforts would exceed the marginal benefit.

KEY TAKEAWAYS

- Pollution is related to the concept of scarcity. The existence of pollution implies that an environmental resource has alternative uses and is thus scarce.
- Pollution has benefits as well as costs; the emission of pollutants benefits people by allowing other activities to be pursued at lower costs. The efficient rate of emissions occurs where the marginal benefit of emissions equals the marginal cost they impose.
- The marginal benefit curve for emitting pollutants can also be read from right to left as the marginal cost of abating emissions. The marginal cost curve for increased emission levels can also be read from right to left as the demand curve for improved environmental quality.
- The Coase theorem suggests that if property rights are well-defined and if transactions are costless, then the private market will reach an efficient solution. These conditions, however, are not likely to be present in typical environmental situations. Even if such conditions do not exist, Coase's arguments still yield useful insights to the mitigation of environmental problems.
- Surveys are sometimes used to measure the marginal benefit curves for emissions and the marginal cost curves for increased pollution levels. Marginal cost curves may also be inferred from other relationships. Two that are commonly used are the demand for housing and the relationship between pollution and production.

TRY IT!

The table shows the marginal benefit to a paper mill of polluting a river and the marginal cost to residents who live downstream. In this problem assume that the marginal benefits and marginal costs are measured at (not between) the specific quantities shown.

Plot the marginal benefit and marginal cost curves. What is the efficient quantity of pollution? Explain why neither one ton nor five tons is an efficient quantity of pollution. In the absence of pollution fees or taxes, how many units of pollution do you expect the paper mill will choose to produce? Why?

Quantity of pollution (tons per week)	Marginal benefit	Marginal cost
0	$110	$0
1	100	8
2	90	20
3	80	35
4	70	70
5	60	150
6	0	300

Case in Point: Estimating a Demand Curve for Environmental Quality

How do economists estimate demand curves for environmental quality? One recent example comes from work by Louisiana State University economist David M. Brasington and Auburn University economist Diane Hite. Using data from Ohio's six major metropolitan areas (Akron, Cincinnati, Cleveland, Columbus, Dayton, and Toledo), the economists studied the relationship between house prices and the distance between individual houses and hazardous waste sites. From this, they were able to estimate the demand curve for environmental quality—at least in terms of the demand for locations farther from environmental hazards.

The economists used actual real estate transactions in 1991 to get data on house prices. In that year, there were 1,192 hazardous sites in the six metropolitan areas. The study was based on 44,255 houses. The median distance between a house and a hazardous site was 1.08 miles. The two economists found, as one would expect, that house prices were higher the greater the distance between the house and a hazardous site. All other variables unchanged, increasing the distance from a house to a hazardous site by 10% increased house value by 0.3%.

Other characteristics of the demand curve shed light on the relationship between environmental quality and other goods. For example, the study showed that people substitute house size for environmental quality. A house closer to a hazardous waste site is cheaper; people take advantage of the lower price of such sites to purchase larger houses.

While house size and environmental quality were substitutes, school quality, measured by student scores on achievement tests, was a complement. If the price of school quality were to fall by 10%, household would buy 8% more environmental quality. The cross-price elasticity between environmental quality and school quality was thus estimated to be −0.80.

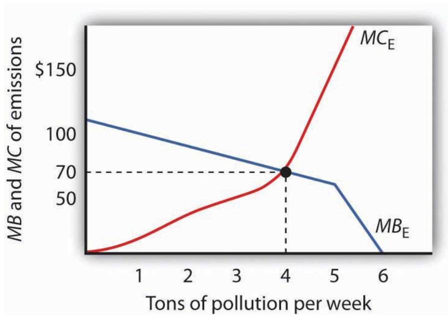
There is no marketplace for environmental quality. Estimating the demand curve for such quality requires economists to examine other data to try to infer what the demand curve is. The study by Professors Brasington and Hite illustrates the use of an examination of an actual market, the market for houses, to determine characteristics of the demand curve for environmental quality.

Source: David M. Brasington and Diane Hite, "Demand for Environmental Quality: A Spatial Hedonic Analysis," Regional Science & Urban Economics, 35(1) (January 2005): 57–82.

ANSWER TO TRY IT! PROBLEM

The efficient quantity of pollution is four tons per week. At one ton of pollution, the marginal benefit exceeds the marginal cost. If the paper mill expands production, the additional pollution generated leads to additional benefits for it that are greater than the additional cost to the residents nearby. At five tons, the marginal cost of polluting exceeds the marginal benefit. Reducing production, and hence pollution, brings the marginal costs and benefits closer.

In the absence of any fees, taxes, or other charges for its pollution, the paper mill will likely choose to generate six tons per week, where the marginal benefit has fallen to zero.



2. ALTERNATIVES IN POLLUTION CONTROL

LEARNING OBJECTIVES

1. **Explain why command-and-control approaches to reducing pollution are inefficient.**
2. **Explain what is meant by incentive approaches to pollution reduction and discuss examples of them, explaining their advantages and disadvantages.**

Suppose that the goal of public policy is to reduce the level of emissions. Three types of policy alternatives could be applied. The first is persuasion—people can be exhorted to reduce their emissions. The second relies on direct controls, and the third uses incentives to induce reductions in emissions.

2.1 Moral Suasion

Back in the days when preventing forest fires was a goal of public policy, a cartoon character called Smokey the Bear once asked us to be careful with fire. Signs everywhere remind us not to litter. Some communities have mounted campaigns that admonish us to "Give a hoot—don't pollute."

These efforts to influence our choices are examples of **moral suasion**, an effort to change people's behavior by appealing to their sense of moral values. Moral suasion is widely used as a tactic in attempting to reduce pollution. It has, for example, been fairly successful in campaigns against littering, which can be considered a kind of pollution. Efforts at moral suasion are not generally successful in reducing activities that pollute the air and water. Pleas that people refrain from driving on certain days when pollution is very great, for example, achieve virtually no compliance.

moral suasion

An effort to change people's behavior by appealing to their sense of moral values.

Moral suasion appears to be most effective in altering behaviors that are not already widespread and for which the cost of compliance is low. It is easy to be careful with one's fires or to avoid littering.

Moral suasion does not, however, appear to lead to significant changes in behavior when compliance costs are high or when the activity is already widely practiced. It is therefore not likely to be an effective tool for reducing most forms of pollution.

2.2 Command and Control

command-and-control approach

Approach in which a government agency specifies how much or by what method a polluting agent must adjust its emissions.

In the most widely used regulatory approach to environmental pollution, a government agency tells a polluting agent how much pollution it can emit or requires the agent to use a particular production method aimed at reducing emissions. This method, in which the government agency tells a firm or individual how much or by what method emissions must be adjusted, is called the **command-and-control-approach**.

Economists are generally critical of the command-and-control approach for two reasons. First, it achieves a given level of emissions reduction at a higher cost than what would be required to achieve that amount of reduction if market incentives (discussed below) were implemented. Second, it gives polluters no incentive to explore technological and other changes that might reduce the demand for emissions.

Suppose two firms, A and B, each dump 500 tons of a certain pollutant per month and that there is no fee imposed (that is, the price for their emissions equals zero). Total emissions for the two firms thus equal 1,000 tons per month. The pollution control authority decides to cut this in half and orders each firm to reduce its emissions to 250 tons per month, for a total reduction of 500 tons. This is a command-and-control regulation because it specifies the amount of reduction each firm must make. Although it may seem fair to require equal reductions by the two firms, this approach is likely to generate excessive costs.

Suppose that Firm A is quite old and that the reduction in emissions to 250 tons per period would be extremely costly. Suppose that removing the 251st ton costs this firm $1,000 per month. Put another way, the marginal benefit to Firm A of emitting the 251st ton would be $1,000.

Suppose Firm B, a much newer firm, already has some pollution-control equipment in place. Reducing its emissions to 250 tons imposes a cost, but a much lower cost than to Firm A. Indeed, suppose Firm B could reduce its emissions to 249 tons at an additional cost of $100; the marginal benefit to Firm B of emitting the 249th ton is $100.

If two firms have different marginal benefits of emissions, the allocation of resources is inefficient. The same level of emissions could be achieved at a lower cost. Suppose, for example, Firm A is permitted to increase its emissions to 251 tons while Firm B reduces emissions to 249. Firm A saves $1,000, while the cost to Firm B is just $100. Society achieves a net gain of $900, and the level of emissions remains at 500 tons per month.

least-cost reduction in emissions

Situation in which emissions are reduced so that the marginal benefit of an additional unit of pollution is the same for all polluters.

As long as Firm A's marginal benefit of emissions exceeds Firm B's, a saving is realized by shifting emissions from B to A. At the point at which their marginal benefits are equal, no further reduction in the cost of achieving a given level of emissions is possible, and the allocation of emissions is efficient. When a given reduction in emissions is achieved so that the marginal benefit of an additional unit of emissions is the same for all polluters, it is a **least-cost reduction in emissions**. A command-and-control approach is unlikely to achieve a least-cost reduction in emissions.

The inefficiency of command-and-control regulation is important for two reasons. First, of course, it wastes scarce resources. If the same level of air or water quality could be achieved at a far lower cost, then surely it makes sense to use the cheaper method. Perhaps even more significant, reliance on command-and-control regulation makes environmental quality far more expensive than it needs to be—and that is likely to result in an unwillingness to achieve the improvements that would be economically efficient.

There is a further difficulty with the command-and-control approach. Once firms A and B have been told to reduce their emissions to 250 tons per period, neither firm has an incentive to try to do better. Neither firm will engage in research seeking to reduce its emissions below 250 tons.

Another way to see the difficulty with command-and-control approaches to pollution control is to imagine that a similar method were in use for other resource allocation problems. Suppose, for example, that labor were allocated according to a command-and-control mechanism. Rather than leaving labor allocation to the marketplace, suppose that firms were simply told how much labor to use. Such a system would clearly be unworkable.

2.3 Incentive Approaches

incentive approaches

Approaches to pollution regulation that create market-like incentives to encourage reductions in pollution but allow individual decision makers to decide how much to pollute.

Markets allocate resources efficiently when the price system confronts decision makers with the costs and benefits of their decisions. Prices create incentives—they give producers an incentive to produce more and consumers an incentive to economize. Regulatory efforts that seek to create market-like incentives to encourage reductions in pollution, but that allow individual decision makers to determine how much to pollute, are called **incentive approaches**.

Emissions Taxes

One incentive approach to pollution control relies on taxes. If a tax is imposed on each unit of emissions, polluters will reduce their emissions until the marginal benefit of emissions equals the tax, and a least-cost reduction in emissions is achieved.

Emissions taxes are widely used in Europe. France, for example, has enacted taxes on the sulfur dioxide and nitrous oxide emissions of power plants and other industrial firms. Spain has imposed taxes on the dumping of pollutants into the country's waterways. Emissions taxes have long been imposed on firms that dump pollutants into some river systems in Europe.

Emissions taxes are also being used in economies making the transition from socialism to a market system. In China, taxes are used to limit the emission by firms of water pollutants. By law, 80% of the money collected from these taxes goes back to the firms themselves for emissions control projects. The tax in China is a rudimentary one. The level of a firm's emissions is determined by visual inspection of the water just downstream from the point at which the firm emits pollutants. Taxes are imposed based on a guess as to how much the firm has emitted. A similar approach is used in China to limit emissions of particulate matter, an important air pollutant. While the method appears primitive, it is nevertheless successful. During the 1990s, China had the fastest growing economy in the world. Total economic activity increased in China at an annual rate of about 10% per year. Despite this phenomenal increase in economic activity, international estimates of China's emissions of particulates suggest that they were down 50% during the decade. China's emissions of effluents were unchanged during this period of phenomenal growth.

Taxes are also used to limit pollution in the countries of the former Soviet Union. Rather than basing taxes on an estimate of marginal cost, taxes are set according to complicated engineering formulae. In Lithuania, for example, the tax on sulfide emissions in the water has been set at several million dollars per ton. Needless to say, that tax is not even collected.

The important role for pollution taxes in improving environmental quality is often misunderstood. For example, an intriguing battle in the courts followed Argentina's attempt to use taxes to control air pollution. Environmental groups went to federal court, charging that the taxes constituted a "license to pollute." The unfortunate result was that the Argentine government withdrew its effort to control pollution through taxation without finding an adequate substitute policy.[5]

An emissions tax requires, of course, that a polluter's emissions be monitored. The polluter is then charged a tax equal to the tax per unit times the quantity of emissions. The tax clearly gives the polluter an incentive to reduce emissions. It also ensures that reductions will be accomplished by those polluters that can achieve them at the lowest cost. Polluters for whom reductions are most costly will generally find it cheaper to pay the emissions tax.

In cases where it is difficult to monitor emissions, a tax could be imposed indirectly. Consider, for example, farmers' use of fertilizers and pesticides. Rain may wash some of these materials into local rivers, polluting the water. Clearly, it would be virtually impossible to monitor this runoff and assess responsible farmers a charge for their emissions. But it would be possible to levy a tax on these materials when they are sold, confronting farmers with a rough measure of the cost of the pollution their use of these materials imposes.

Marketable Pollution Permits

An alternative to emissions taxes is marketable pollution permits, which allow their owners to emit a certain quantity of pollution during a particular period. The introduction to this chapter dealt with an example of marketable pollution permits; each of the permits to dump a metric ton of greenhouse gases that is purchased results in a reduction in the cost of meeting the goal of reducing the emissions of these gases.

To see how this works, suppose that Firms A and B are again told that they must reduce their emissions to 250 tons of a pollutant per month. This time, however, they are given 250 permits each—one permit allows the emission of one ton per month. They can trade their permits; a firm that emits less than its allotted 250 tons can sell some of its permits to another firm that wants to emit more.

We saw that Firm B can reduce its emissions below 250 tons for a much lower cost than Firm A. For example, it could reduce its emissions to 249 tons for $100. Firm A would be willing to pay $1,000 for the right to emit the 251st ton of emissions. Clearly, a mutually profitable exchange is possible. In fact, as long as their marginal benefits of pollution differ, the firms can profit from exchange. Equilibrium will be achieved at the least-cost solution at which the marginal benefits for both firms are equal.

One virtue of using marketable permits is that this approach represents only a modest departure from traditional command-and-control regulation. Once a polluter has been told to reduce its emissions to a certain quantity, it has a right to emit that quantity. Polluters will exchange rights only if doing so increases their utility or profits—allowing rights to be exchanged can only make them better off. Another benefit, of course, is that such exchanges allow a shift from the inefficient allocation created by command-and-control regulation to an efficient allocation in which pollution is reduced at the lowest possible cost. Finally, each firm will have an incentive to explore ways to reduce its emissions further, either to be able to sell more rights or to require purchasing fewer permits.

Merits of Incentive Approaches

Incentive systems, either emissions taxes or tradable permits, can achieve reductions in emissions at a far lower cost than command-and-control regulation. Even more important, however, are the long-run incentives they create for technological change. A firm that is ordered to reduce its emissions to a certain level has no incentive to do better, whereas a firm facing an emissions tax has a constant incentive to seek out new ways to lower its emissions and thus lower its taxes. Similarly, a firm faces a cost for using an emissions permit—the price that could be obtained from selling the permit—so it will seek ways to reduce emissions. As firms discover new ways to lower their costs of reducing emissions, the demand for emissions permits will fall, lowering the efficient quantity of emissions—and improving environmental quality even further.

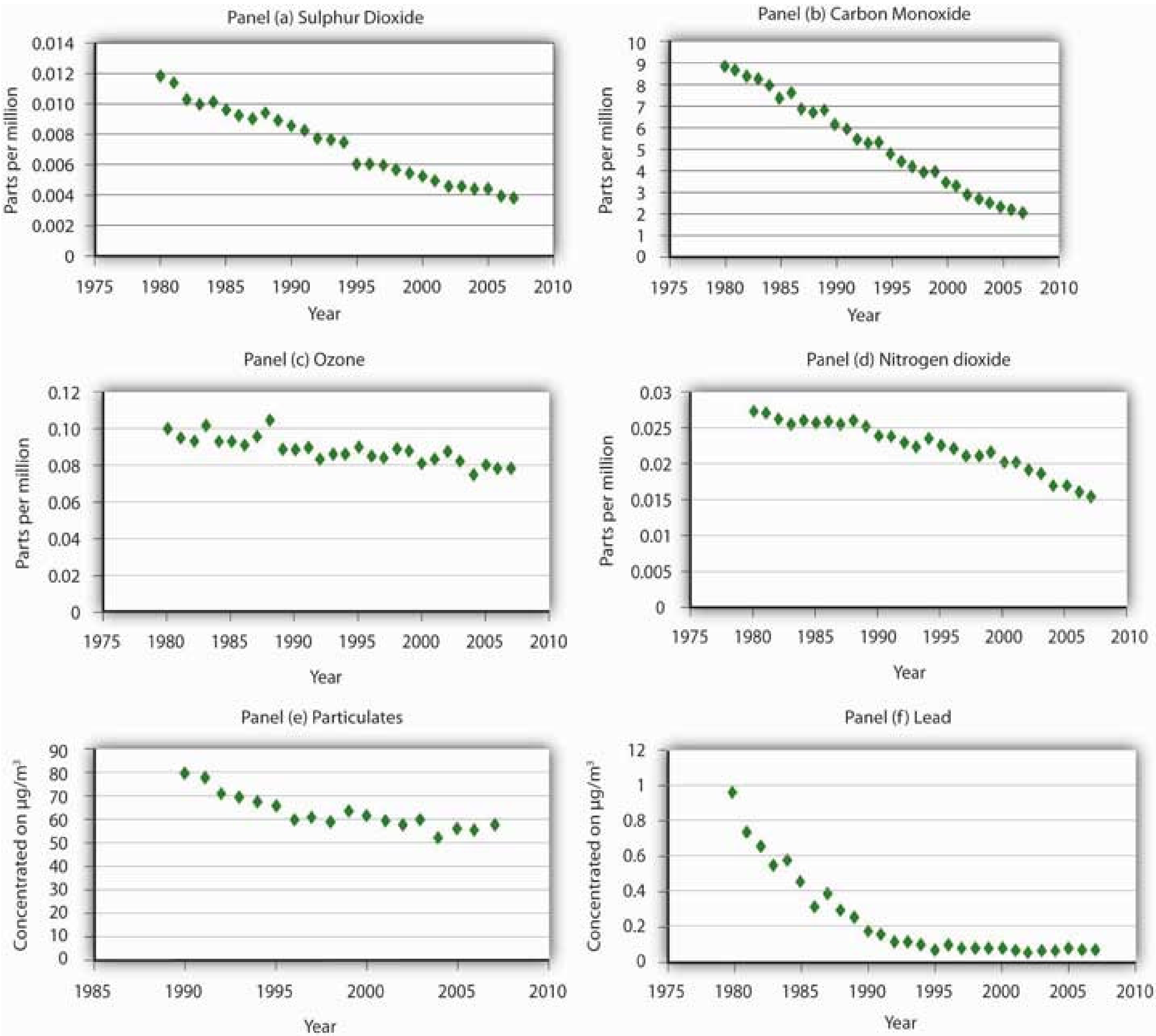
2.4 Public Policy and Pollution: The Record

Federal efforts to control air and water pollution in the United States have produced mixed results. Air quality has generally improved; water quality has improved in some respects but deteriorated in others.

Figure 18.7 shows how annual average concentrations of airborne pollutants in major U.S. cities have declined since 1975. Lead concentrations have dropped most dramatically, largely because of the increased use of unleaded gasoline.

FIGURE 18.7 **U.S. Air Pollution Levels, 1975–2003**

The average concentration of major air pollutants in U.S. cities has declined since 1975.



Source: U.S. Bureau of the Census, Statistical Abstract of the United States, 2006 (Washington, D.C.: U.S. Government Printing Office, 2006), Table 359, p. 228. Observations before 1999 are taken from previous volumes of the abstract; supplemented by National Air Quality and Emissions Trends Report, 2003, Table A-1b (Environmental Protection Agency, 2005).

Public policy has generally stressed command-and-control approaches to air and water pollution. To reduce air pollution, the EPA sets air quality standards that regions must achieve, then tells polluters what adjustments they must make in order to meet the standards. For water pollution, the Environmental Protection Agency (EPA) has set emissions limits based on the technologies it considers reasonable to require of polluters. If the implementation of a particular technology will reduce a polluter's emissions by 20%, for example, the EPA will require a 20% reduction in emissions. National standards have been imposed; no effort has been made to consider the benefits and costs of pollution in individual streams. Further, the EPA's technology-based approach pays little attention to actual water quality—and has produced few gains.

Moreover, environmental problems go beyond national borders. For example, sulfur dioxide emitted from plants in the United States can result in acid rain in Canada and elsewhere. Another possible pollution problem that extends across national borders is suggested by the global warming hypothesis.

Many scientists have argued that increasing emissions of greenhouse gases, caused partly by the burning of fossil fuels, trap ever more of the sun's energy and make the planet warmer. Global warming, they argue, could lead to flooding of coastal areas, losses in agricultural production, and the elimination of many species. If this global warming hypothesis is correct, then carbon dioxide is a pollutant with global implications, one that calls for a global solution.

At the 1997 United Nations conference in Kyoto, Japan, the industrialized countries agreed to cuts in carbon dioxide emissions of 5.2% below the 1990 level by 2010. At the 1998 conference in Buenos Aires, Argentina, 170 countries agreed on a two-year action plan for designing mechanisms to reduce emissions and procedures to encourage transfers of new energy technologies to developing countries.

While the delegates to the conferences sign the various protocols, countries are not bound by them until ratified by appropriate governmental bodies. By 2005, the vast majority of the world's nations had ratified the agreement. The United States had neither signed nor ratified the agreement.

Debates at these conferences have been over the extent to which developing countries should be required to reduce their emissions and over the role that market mechanisms should play. Developing countries argue that their share in total emissions is fairly small and that it is unfair to ask them to give up economic growth, given their lower income levels. As for how emissions reductions will be achieved, the United States has been the strongest advocate for emissions trading in which each country would receive a certain number of tradable emissions rights. This provision has been incorporated in the agreement.

That market approaches have entered the national and international debates on dealing with environmental issues, and to a large extent have even been used, demonstrates the power of economic analysis. Economists have long argued that as pollution-control authorities replace command-and-control strategies with incentive approaches, society will reap huge savings. The economic argument has rested on acknowledging the opportunity costs of addressing environmental concerns and thus has advocated policies that achieve improvements in the environment in the least costly ways.

KEY TAKEAWAYS

- Public sector intervention is generally needed to move the economy toward the efficient solution in pollution problems.
- Command-and-control approaches are the most widely used methods of public sector intervention, but they are inefficient.
- The exchange of pollution rights can achieve a given reduction in emissions at the lowest cost possible. It also creates incentives to reduce pollution demand through technological change.
- Tax policy can also achieve a least-cost reduction in emissions.

TRY IT!

Based on your answer to the previous Try It! problem, a tax of what amount would result in the efficient quantity of pollution?

Case in Point: Road Pricing in Singapore

The urban highways of virtually every city in the world are heavily congested in what is fancifully referred to as the "rush hour." Traffic during this period hardly rushes. The problem of traffic congestion is analogous to the problem of pollution, and it lends itself to the same solution.

Suppose that you are driving into any major city on a weekday at eight o'clock in the morning. The highway is congested when you drive onto it. Your car adds to that congestion. You, of course, experience the average level of congestion. But, your car slows every car behind you on the highway by a small amount. Multiplying that extra slowing by the number of cars behind you gives the marginal delay of adding your own car to an already congested highway. That marginal cost is many times greater than the average cost that you actually face. The result is an inefficient solution in which roads are congested far beyond the point that would be economically efficient. The late William Vickrey, who won the Nobel Prize for his work in the economics of public finance, advocated a system of road pricing that would put an end to congested highways.

Mr. Vickrey's dream is a reality in Singapore. The island-nation's 700,000 cars are each required to subscribe to the Electronic Road Pricing (ERP) system and to have an ERP card on their windshield. Further, each driver is required to maintain a deposit of electronic cash in the card. Tolls are charged during rush hour for the highways leading into the downtown area. There is also a charge for using downtown streets. Tolls range from $0.50 to $1.00 and are only levied if the road threatens to become congested. Using highways and streets is free at other times. ERP managers adjust the tolls to avoid "jam-ups," as they are called in Singapore. Managers attempt to keep traffic flowing at 30 to 40 mph on the highways and 15 to 20 mph on downtown streets.

When a car passes an electronic gurney, it automatically takes the charge out of the ERP card on the windshield. When a charge is taken from the card, the driver hears a beep. If the card balance falls below $5, several beeps are heard. The cards can be placed in ATMs to put additional funds in them. A car that does not have enough money in its card is photographed and the owner of the car receives a $45 citation in the mail. The tolls keep casual drivers off the roads during rush hour. Commuters generally find it cheaper to use the island's excellent rapid transit system.

The system is not universally popular. People in Singapore refer to the system as "Eternally Raising Prices." The manager of the system, Hiok Seng Tan, rejects the criticism. "It's not just a system to get more money of Singaporeans," he told the *Boston Globe*. In fact, the system takes in only $44 million per year; the money is used to maintain the system.

Singapore's system is an example of a price system to manage what is, in effect, a pollution problem. It appears to be effective. Jam-ups are uncommon. Would such a system work in other areas? The idea is not politically popular. Gregory B. Christainsen, a professor of economics at California State University at East Bay, notes that Singapore is dominated by a single party and questions whether other urban areas are ready for the approach. Still, the system works and the concept is one more urban areas should consider.

Sources: Alex Beam, "Where Traffic Has Been Tamed," Boston Globe, (May 2, 2004): M-7; Gregory B. Christainsen, "Road Pricing in Singapore after 30 Years," Cato Journal, 26(1) (Winter 2006): 71–88.

ANSWER TO TRY IT! PROBLEM

The paper mill will reduce its pollution until the marginal benefit of polluting equals the tax. In this case a tax of $70 would cause the paper mill to reduce its pollution to 4 tons, the efficient level. At pollution levels below that amount, the marginal benefit of polluting exceeds the tax, and so the paper mill is better off polluting and paying the tax. At pollution levels greater than that amount, the marginal benefit of polluting is less than the tax.

3. REVIEW AND PRACTICE

Summary

Pollution is a by-product of human activity. It occurs when the environment becomes scarce—when dumping garbage imposes a cost. There are benefits as well as costs to pollution; the efficient quantity of pollution occurs where the difference between total benefits and total costs is maximized. This solution is achieved where the marginal benefit of additional pollution equals the marginal cost. We have seen that an alternative approach shows that efficiency is also achieved where the marginal benefit of pollution abatement equals the marginal cost of abatement.

Economists measure the benefits of pollution in terms of the costs of not dumping the pollution. The same curve can be read from left to right as the marginal benefit curve for emissions and from right to left as the marginal cost curve for abatement.

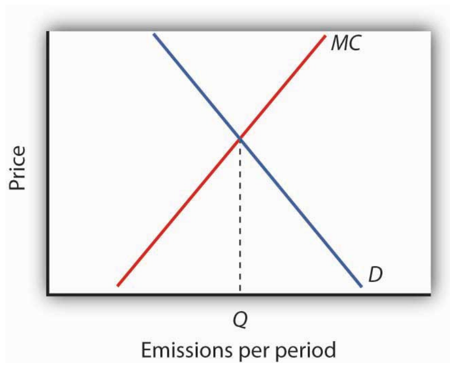
The costs of pollution are measured in two ways. One is through direct surveys. Respondents can be asked how much compensation they would be willing to accept in exchange for a reduction in environmental pollution; alternatively, they can be asked how much they would pay for an improvement in environmental quality. A second approach infers the marginal cost of increased pollution from other relationships. The effects of pollution on house prices or rental values, for example, allow economists to estimate the value people place on environmental quality. Pollution costs can also be estimated on the basis of the costs they impose on firms in production.

Three types of policies are available to reduce pollution. Moral suasion is sometimes used, but it is effective only under limited conditions. Command-and-control regulation is used most commonly, but it is likely to be inefficient. It also fails to provide incentives for technological change in the long run. The most promising policies are the incentive approaches, which include emissions taxes and marketable pollution permits. Both can be designed to reduce emissions at the lowest cost possible, and both create an incentive for firms to search out new and cheaper ways to reduce emissions.

Although public policy has stressed command-and-control methods in the past, pollution rights exchanges are now being introduced. Past policies may have been inefficient, but they have succeeded in improving air quality, at least in the nation's cities.

CONCEPT PROBLEMS

1. We have noted that economists consider the benefits and costs of pollution from the perspective of people's preferences. Some critics argue, however, that the interests of plants and animals should be considered: for example, if pollution is harming trees, the trees have a right to protection. Do you think that is a good idea? How would it be implemented?
2. List five choices you make that result in pollution. What price do you pay to pollute the environment? Does that price affect your choices?
3. In any urban area, what group is likely to be exposed to a greater level of pollution—rich people or poor people? (*Hint:* Utilize the findings of economists concerning the relationship between house prices and pollution levels.)
4. Suppose the accompanying graph shows the demand and marginal cost curves, *D* and *MC*, for a pollutant in a particular area. How do you think future economic and population growth will affect the efficient rate of emissions per period, *Q*, and thus the level of pollution?



5. "Environmental quality is not just a matter of technical efficiency; it's about how people relate to nature. Economists are completely off base in their analysis of the benefits and costs of pollution." What is your opinion of this quote?
6. Campaigns that exhort us to "Give a hoot—don't pollute" imply that anyone who pollutes the environment is uncaring—that people who are concerned about environmental quality would not be dumping garbage into the environment. Is that true?
7. We have seen that a system of marketable pollution permits achieves the same solution as a system of emissions taxes. Which do you think would be fairer? Why?
8. Many environmentalists are horrified by the notion of marketable pollution permits advocated by economists. These environmentalists insist that pollution is wrong, and that no one should be able to buy the right to pollute the environment. What do you think?
9. Some people object that charging firms for their emissions will do no good—firms will simply raise their prices and go on doing what they were doing before. Comment on this objection to emissions taxes.
10. Suppose firms in a perfectly competitive industry generate water pollution as a by-product of their production, and they are not charged for this. Who benefits from their use of the environment as a dumping ground? If an emissions tax is imposed, costs to these firms increase and emissions drop. Who will bear the burden of this tax? Is that fair?
11. The Case in Point on measurement suggested that the demand curve for locating a house farther from a hazardous waste site could be inferred from property value studies. Explain how this could be the case.
12. Does the road pricing system in Singapore strike you as fair? Would you like to see such a system in your own area? Why or why not?

NUMERICAL PROBLEMS

1. Suppose the dry-cleaning industry is perfectly competitive. The process of dry cleaning generates emissions that pollute the air, and firms now emit this pollution at no cost. Suppose that the long run equilibrium price for dry cleaning a typical item is $5, and a pollution-control program increases the marginal cost by $1 per item.
    a. How will the pollution-control program affect the price of dry-cleaning services in the short run? Explain and illustrate graphically.
    b. Explain and illustrate graphically how the program will affect the output of dry-cleaning services in the short run.
    c. Explain and illustrate graphically how the $1 increase in cost will affect the price of dry-cleaning services in the long run.
    d. Explain and illustrate graphically how the cost increase will affect the quantity of dry-cleaning services produced in the long run.
    e. Who pays for pollution control in the industry? Explain, relating your answer to the basic conclusion about long-run equilibrium in a perfectly competitive industry.
2. Now suppose the dry-cleaning industry in the community is monopolistically competitive. Suppose the initial price per unit of dry-cleaning is $6. Suppose that a charge levied on dry-cleaning firms for the pollution they generate increases the cost of a unit of dry-cleaning by $1.
    a. Explain and illustrate graphically how the $1 charge will affect the price charged by typical firm in the short run.
    b. Explain and illustrate graphically how the $1 charge will affect the typical firm's output in the short run.
    c. Now explain and illustrate graphically how the $1 charge will affect price and output of a typical firm in the long run. Through what mechanism does this occur?
    d. Compare your answers for a world of monopolistically competitive firms to a world of perfectly competitive firms. Is there any significant difference between the conclusions of the two models?
3. Suppose local government regulations allow only a single firm to provide dry-cleaning services to a local community, and this firm generates pollution as in Problem 1. The firm initially charges a price of $4 per item. Now a pollution-control program is imposed, increasing the firm's marginal and average total costs by $1 per item.
    a. Explain and illustrate graphically how the program will affect the firm's price and output.
    b. Who pays for the pollution-control program?
4. Suppose the marginal benefit (*MB*) and marginal cost (*MC*) curves for emitting particulate matter are given by the following schedules, where *E* is the quantity of emissions per period. The marginal benefits and costs are measured at the quantities of emissions shown.

E/period	MB	MC
0	$230	$0
200	190	10
400	150	30
600	110	50
800	70	70
1,000	30	90

    a. Plot the marginal benefit and marginal cost curves and state the efficient quantity of emissions per period.
    b. What quantity of emissions will occur when the price of emissions is zero?
    c. What tax rate would achieve the efficient rate of emissions?
5. Now suppose that rising incomes increase marginal cost as follows:

E/period	New MC
0	$0
200	30
400	70
600	110
800	150
1,000	190

a. Plot the new marginal cost curve in the graph you drew in Problem 4. What is the new efficient quantity of emissions per period?
b. What quantity of emissions will occur when the price of emissions is zero?
c. What tax rate would achieve the efficient rate of emissions?

6. The text contains the following statement: "All other variables unchanged, increasing the distance from a house to a hazardous site by 10% increased house value by 0.3%." What is the price elasticity of house prices with respect to proximity to a hazardous waste site?

ENDNOTES

1. *ECX Market Update*, June 2008 (available at europeanclimateexchange.com).
2. "A Green Future," *Economist*, 372(8392) (September 11, 2004): 69–70; Jeffrey Ball, "New Market Shows Industry Moving on Global Warming," *The Wall Street Journal*, 241(11) (January 16, 2003): A1; John J. Fialka, "Russian Interest in Signing Kyoto Spurs Trading," *The Wall Street Journal Online*, June 1, 2004.
3. R. H. Coase, "The Problem of Social Cost," *Journal of Law and Economics* 3 (October 1960): 1–44.
4. See, for example, Nir Becker and Doron Lavee, "The Benefits and Costs of Noise Reduction," *Journal of Environmental Planning and Management*, 46(1) (January 2003): 97–111, which shows the negative relationship between apartment prices and noise levels (a form of pollution) in Israel.
5. The observations on pollution taxes in China, Argentina, and Lithuania are from Randall A. Bluffstone, "Environmental Taxes in Developing and Transition Economies," *Public Finance & Management* 3(1) (Spring, 2003): 143–75.