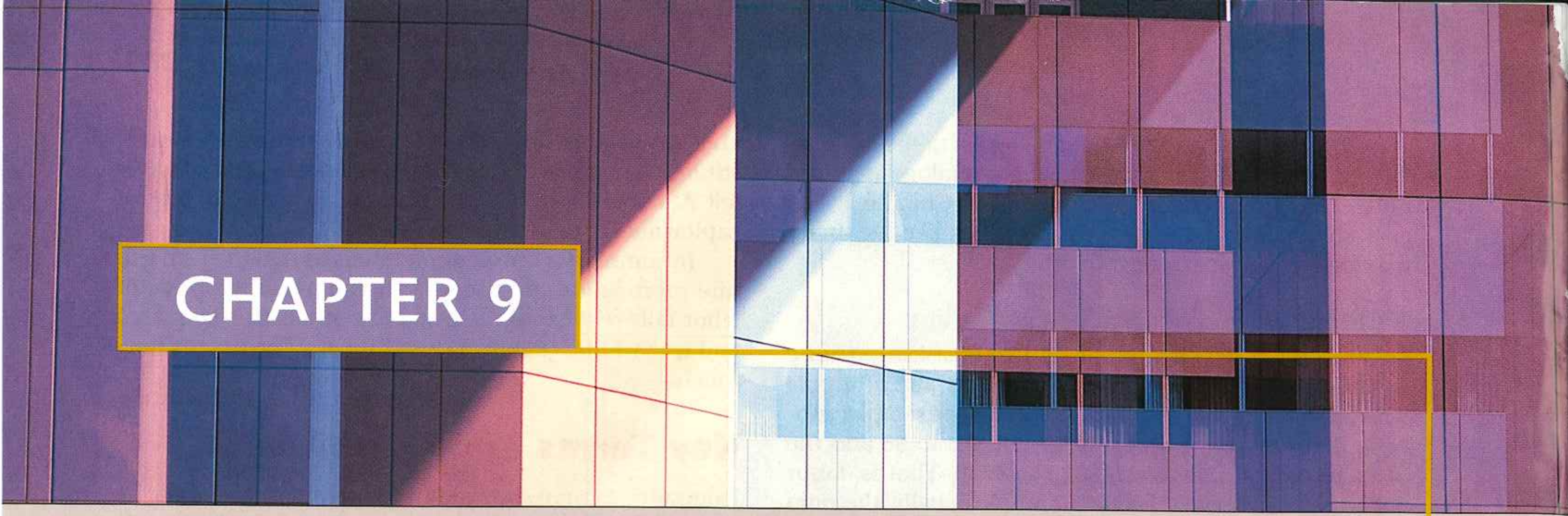


CHAPTER 9

The Rise and Fall of Industries

Fred Smith's college term paper led to the birth of a new industry. In the paper, he described his idea for a new product: reliable overnight mail service. Although he got only a C on the paper, Fred Smith pursued his idea. He became an entrepreneur. After college, in 1973, he started a business firm that guaranteed next-day delivery of a letter or a package virtually anywhere in the United States. The firm, Federal Express, was successful, very successful; its sales reached $1 billion by 1982, $4 billion by 1988, $8 billion by 1992, and over $29 billion by 2005.

Seeing high profits at Federal Express, many other firms entered the express delivery industry. In the late 1970s, United Parcel Service (UPS) entered; in the early 1980s, the U.S. Postal Service entered; many small local firms you've probably never heard of also got into the act. The entire industry expanded along with Federal Express.

The express delivery industry is an example of an industry on the rise. Many other examples of fast-growing industries exist in the annals of economic history. Estée Lauder founded a cosmetic firm 50 years ago; it grew along with the cosmetics industry as a whole.

Kemmons Wilson started the motel franchising industry when he saw the potential demand for clean, reliable rooms for travelers and opened his first Holiday Inn in Memphis in 1952; by 1968, there were 1,000 Holiday Inns, and now the industry includes other motel firms such as Days Inn and Motel 6.

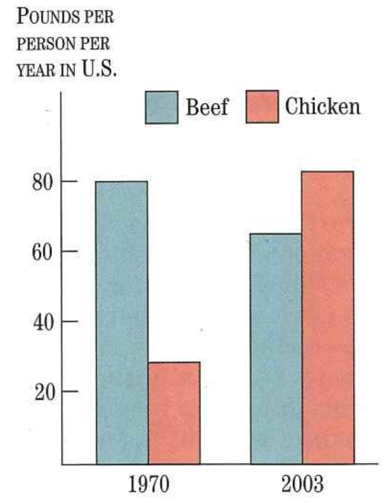


Figure 9.1
Taste Shifts and the Rise and Fall of Different Industries
Changes in tastes cause some industries to grow and others to contract. Concerns about fat in the diet may have been one reason that consumer tastes shifted from beef to chicken in the United States. In any case, the chicken industry has flourished and the beef industry has suffered as a result of the taste shift. Can you think of any other reasons why the per capita consumption of chicken increased in this period? Given your knowledge of supply and demand, how would you test your hypothesis?

Of course, industries do not always grow. The U.S. beef industry has declined as the U.S. chicken industry has risen. The mainframe computer industry has declined as the personal computer industry has risen.

The causes of the rise and fall of industries can be traced to new ideas such as overnight delivery, to new cost-reducing technologies such as the Internet, or to changes in consumer tastes, such as a shift in preference toward foods with less fat. This latter shift, for example, is one reason behind the rise of the chicken industry and the fall of the beef industry, as shown in Figure 9.1. In 2003 and 2004, the widespread popularity of low-carb diets, which favor reducing the intake of foods like bread and pasta in favor of high-protein foods like meat and cheese, caused some concern among bread producers and retailers and prompted several major U.S. food companies to produce low-carb versions of many products. Other companies entered the market with low-carb specialty products. Some industries have recurring ups and downs. The oil tanker shipping industry, for example, regularly expands when oil demand increases and declines when oil demand falls.

In this chapter, we develop a model to explain the behavior of whole industries over time. We examine how economic forces cause industries to adjust to new technologies and to shifts in consumer tastes. Our analysis assumes that the firms are operating in competitive markets. The initial forces causing an industry to rise or fall are described by shifts in a cost curve or a demand curve. Changes in the industry then occur as firms either enter or exit the industry. The central task of this chapter is to show how an industry grows or contracts as firms enter or exit the industry. Do profits fall or rise? Do the prices consumers pay increase or decrease? Before addressing these questions, we provide a brief definition of industries and some examples of different industry types.

Markets and Industries

industry: a group of firms producing a similar product.

An **industry** is a group of firms producing a similar product. The cosmetics industry, for example, refers to the firms producing cosmetics. The term *market* is sometimes used instead of industry. For example, the phrases "the firms in the cosmetics industry" and "the firms in the cosmetics market" mean the same thing. But the term *market* can also refer to the consumers who buy the goods and to the interaction of the producers and the consumers. Both firms and consumers are in the cosmetics market, but only firms are in the cosmetics industry.

Manufacturing is the making of goods by mechanical or chemical processes. In economics, the word *industry* is much broader than manufacturing. Firms in an industry can produce *services* such as overnight delivery or overnight accommodations as well as manufactured goods.

Many industries are global. Firms in the United States sell or produce many of their goods in other countries. U.S. firms compete with firms in Japan, Europe, and elsewhere. The aspirin industry has been a global industry for 100 years. Reduced transportation and communication costs in recent years have made most other industries global. Until competition from Europe and Japan intensified 30 years ago, the automobile industry in the United States consisted mainly of three firms—General Motors, Ford, and Chrysler. Now the industry is truly global, with Honda, Toyota, Hyundai, and Nissan selling cars in the United States, and Ford and General Motors selling cars throughout the world.

The Long-Run Competitive Equilibrium Model of an Industry

The model we develop to explain the behavior of industries assumes that firms in the industry maximize profits and that they are competitive. As in the competitive equilibrium model of Chapter 7, individual firms are price-takers; that is, they cannot affect the price. But in order to explain how the industry changes over time, in this chapter we add something new to the competitive equilibrium model: Over time, some firms will enter an industry and other firms will exit an industry. Because the entry and exit of firms takes time, we call this model the **long-run competitive equilibrium model.**

long-run competitive equilibrium model: a model of firms in an industry in which free entry and exit produce an equilibrium such that price equals the minimum of average total cost.

When we use the long-run competitive equilibrium model to explain the behavior of an actual industry, we do not necessarily mean that the industry itself exactly conforms to the assumptions of the model. A model is a means of explaining events in real-world industries; it is not the real world itself. In fact, some industries are very competitive and some are not very competitive. But the model can work well as an approximation in many industries. In Chapters 10 and 11 we will develop alternative models of industry behavior that describe monopoly and the gray area between monopoly and competitive markets. But for this chapter, we focus on the competitive model. This model was one of the first developed by economists to explain the dynamic behavior of an industry; it has wide applicability, and it works well. Moreover, understanding the model will make it easier to understand the alternative models developed in later chapters.

Setting Up the Model with Graphs

The assumption that a competitive firm cannot affect the price is illustrated in Figure 9.2. The left graph views the market from the perspective of a single typical firm in an industry. The price is on the vertical axis, and the quantity produced by the single firm is on the horizontal axis. The market demand curve for the goods produced by

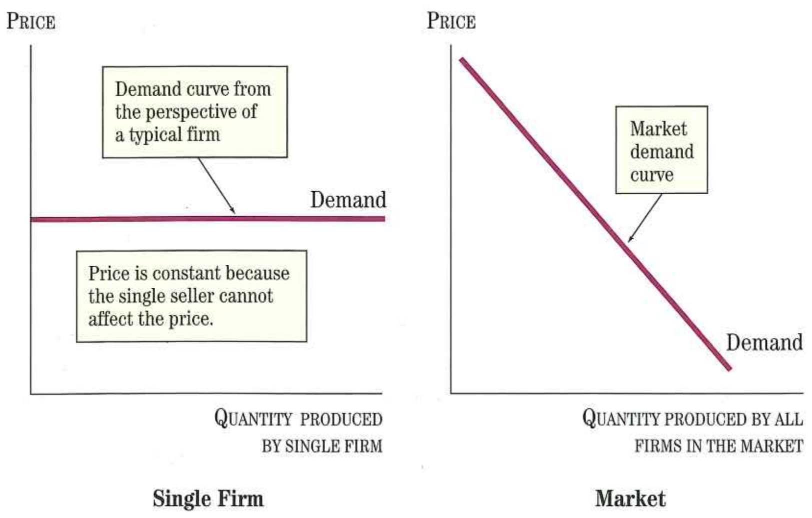


Figure 9.2
How a Competitive Firm Sees Demand in the Market
A competitive market is, by definition, one in which a single firm cannot affect the price. The firm takes the market price as given. Hence, the firm sees a flat demand curve, as shown in the graph on the left. Nevertheless, if all firms change production, the market price changes, as shown in the graph on the right. The two graphs are not alternatives. In a competitive market, they hold simultaneously. (In the graph on the right, a given length along the horizontal axis represents a much greater quantity than the same length in the graph on the left.)

the firms in the industry is shown in the right graph of Figure 9.2. The price is also on the vertical axis of the graph on the right, but the horizontal axis measures the *whole market or industry* production. Because the single firm cannot affect the price, the price, which represents the given market or industry price, is shown by a flat line drawn in the left graph. Notice that even though the single firm takes the price as given, the market demand curve is downward-sloping because it refers to the whole market. If the price in the market rises, then the quantity demanded of the product will fall. If the market price increases, then the quantity demanded will decline.

free entry and exit: movement of firms into and out of an industry that is not blocked by regulation, other firms, or any other barriers.

Entry and Exit. The new characteristic of competitive markets stressed in this chapter is the **free entry and exit** of firms in an industry. The question firms face is whether to *enter* an industry if they are not already in it, or whether to *exit* from an industry they are in. The decisions are based on profits—total revenue less total costs. If profits are positive, there is incentive to enter the industry. If profits are negative, there is incentive to exit the industry. When profits are equal to zero, there is no incentive for either entry or exit.

When firms enter or exit an industry, the entire market or industry supply curve is affected. Recall that the market or industry supply curve is the sum of all the individual firms' supply curves. With more firms supplying goods, the total quantity of goods supplied increases at every price. Thus, more firms in the industry means that the market supply curve shifts to the right; fewer firms in the industry means that the market supply curve shifts to the left.

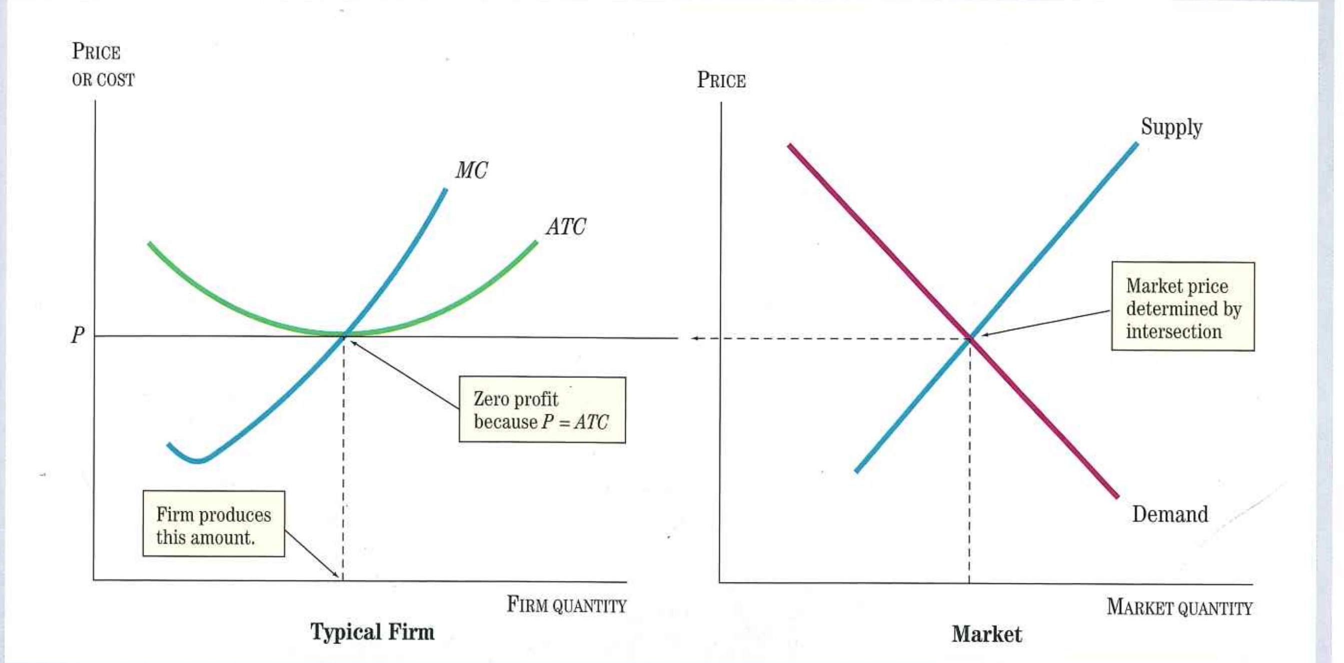


Figure 9.3
Long-Run Equilibrium in a Competitive Market
The left graph shows the typical firm's cost curves and the market price. The right graph shows the market supply and demand curves. The price is the same in both graphs because there is a single price in the market. The price is at a level where profits are zero because price equals average total cost.

Long-Run Equilibrium.

Figure 9.3 is a two-part diagram that shows the profit-maximizing behavior of a typical firm along with the market supply and demand curves. This diagram is generic; it could be drawn to correspond to the numerical specifications of the grape industry or any other industry. In the left graph are the cost curves of the typical firm in the industry with their typical positions: Marginal cost cuts through the average total cost curve at its lowest point. We did not draw in the average variable cost curve in order to keep the diagram from getting too cluttered.

The price line represents the current market price in the industry, for example, the price of a ton of grapes. Because the price line just touches the average total cost curve, we know from Chapter 8 that profits are zero. There is no incentive for firms to either enter or exit the industry. A situation in which profits are zero and there is no incentive to enter or exit—as shown in Figure 9.3—is called a **long-run equilibrium.**

long-run equilibrium: a situation in which entry into and exit from an industry are complete and economic profits are zero, with price (*P*) equal to average total cost (*ATC*).

The market supply and demand curves are to the right of the cost curve diagram in Figure 9.3. The horizontal axis for the market supply and demand curves has a much different scale from that for the individual supply curves. An inch in the right-hand diagram represents much more production than an inch in the left-hand diagram because the diagram on the right is the sum of all the production of all the firms in the market. The market demand curve is downward-sloping: The higher the price, the less the quantity demanded. The intersection of the market supply curve and the market demand curve determines the market price.

The left and right graphs of Figure 9.3 are drawn with the same market price, and this price links the two graphs together. The price touches the bottom of the average total cost curve on the left graph, and this is the price that is at the intersection of the market supply and demand curves. The graphs are set up this way. They are meant to represent a situation of long-run equilibrium: The quantity demanded equals the quantity supplied in the market *and* profits are zero.

An Increase in Demand

Suppose there is a shift in demand—for example, suppose the demand for Zinfandel grapes increases. We show this increase in demand in the top right graph of Figure 9.4; the market demand curve shifts out from *D* to *D*′.

Short-Run Effects. Focus first on the top part of Figure 9.4, representing the short run. With the shift in the demand curve, we move up along the supply curve to a new intersection of the market supply curve and the market demand curve at a higher price. An increase in demand causes a rise in the market price.

Now note in the top left graph that we have moved the price line up from *P* to *P*′. Profit-maximizing firms that are already in the industry will produce more because the market price is higher. This is seen in the top left graph of Figure 9.4; the higher price intersects the marginal cost curve at a higher quantity of production. As production increases, marginal cost rises until it equals the new price.

Note also—and this is crucial—that at this higher price and higher level of production, the typical firm is now earning profits, as shown by the shaded rectangle in the top left graph. Price is above average total cost, and so profits have risen above zero. We have gone from a situation in which profits were zero for firms in the industry to a situation in which profits are positive. Thus, we have moved away from a long-run equilibrium because of the disturbance that shifted the market demand curve. This shift has created a situation in the market in which there is a profit opportunity, encouraging new firms to enter the industry.

Toward a New Long-Run Equilibrium. Now focus on the two graphs in the bottom part of Figure 9.4, representing the long run. They show what happens as new firms enter the industry. In the lower right-hand graph, the supply curve for the whole industry or market shifts to the right from *S* to *S*′. Why? Because the market supply curve is the sum of the individual supply curves, and now there are firms entering the industry and adding to supply.

The rightward shift in the supply curve causes a reduction in the price below *P*′, where it was in the short run. The price will continue adjusting until the price line just touches the bottom of the average total cost curve, where average total cost equals marginal cost. At this point, profits will again be zero and the industry will be in long-run equilibrium. Of course, this adjustment to a new long-run equilibrium takes time. It takes time for firms to decide whether or not to go into business, and it takes time to set up a firm once a decision is made.

The new long-run equilibrium for the typical firm is shown in the lower left graph. It may take several years for an industry to move from the top of Figure 9.4 to the bottom. In fact, it would be more accurate to draw several rows of diagrams between the top and the bottom, showing how the process occurs gradually over time. These additional rows could show more and more firms entering the industry with the price falling until eventually profits are zero again and the incentive to enter the market disappears. The market supply curve will shift to the right until the price comes back to the point where average total cost is at a minimum, where profits are zero, and where no firms will enter or exit the industry.

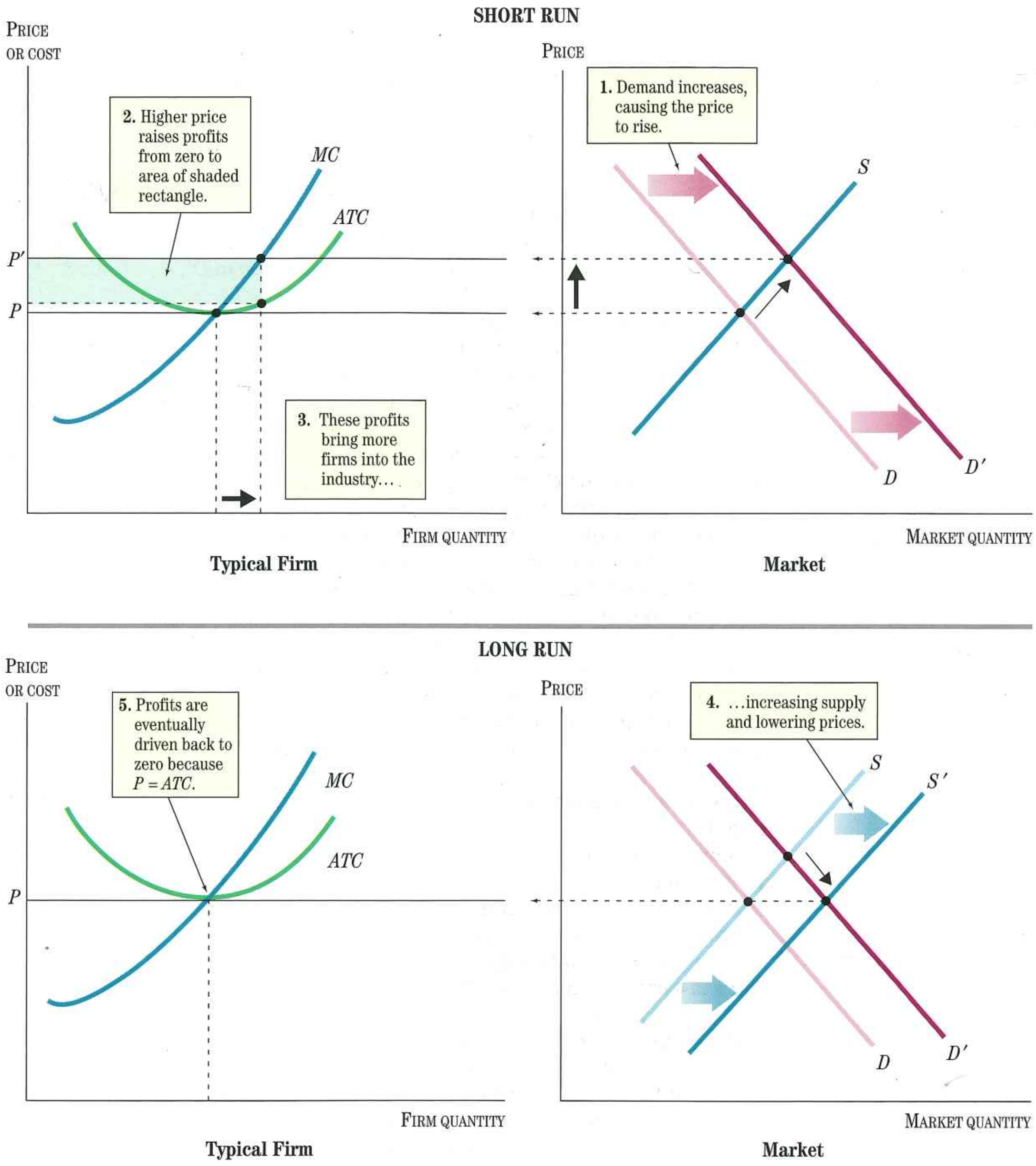

SHORT RUN
PRICE OR COST
2. Higher price raises profits from zero to area of shaded rectangle.
MC
ATC
P′
P
3. These profits bring more firms into the industry. . .
FIRM QUANTITY
Typical Firm
PRICE
1. Demand increases, causing the price to rise.
S
D
D′
MARKET QUANTITY
Market
LONG RUN
PRICE OR COST
5. Profits are eventually driven back to zero because P = ATC.
MC
ATC
P
FIRM QUANTITY
Typical Firm
PRICE
4. . . .increasing supply and lowering prices.
S
S′
D
D′
MARKET QUANTITY
Market

Figure 9.4
The Rise of an Industry after a Shift in Demand
The diagrams at the top show the short run. A shift in the demand curve to the right causes the price to rise from *P* to *P*′; each firm produces more, and profits rise. Higher profits cause firms to enter the industry. The diagrams at the bottom show the long run. As firms enter, the market supply curve shifts to the right, and the price falls back to *P*. New entry does not stop until profits return to zero in the long run.

Economic Profits versus Accounting Profits. It is important at this point to emphasize that the economist's definition of profits is different from an accountant's definition. When you read about the profits of General Motors in the newspaper, it is the accountant's definition that is being reported. There is nothing wrong with the accountant's definition of profits, but it is different from the economist's definition. When an accountant calculates profits for a firm, the total costs do not include the opportunity cost of the owner's time or the owner's funds. Such opportunity costs are *implicit:* The wage that the owner could get elsewhere and the interest that could be earned on the funds if they were invested elsewhere are not explicitly paid, and the accountant therefore ignores them. When computing **accounting profits,** such implicit opportunity costs are *not* included in total costs. When computing **economic profits**—the measure of profits economists use—implicit opportunity costs are included in total costs. Economic profits are equal to accounting profits less any opportunity costs the accountants did not include when measuring total costs.

accounting profits: total revenue minus total costs, where total costs exclude the implicit opportunity costs; this is the definition of profits usually reported by firms.

economic profits: total revenue minus total costs, where total costs include opportunity costs, whether implicit or explicit.

For example, suppose accounting profits for a bakery are $40,000 a year. Suppose the owner of the bakery could earn $35,000 a year working as a manager at a video rental store. Suppose also that the owner could sell the bakery business for $50,000 and invest the money in a bank, where it would earn interest at 6 percent per year, or $3,000. Then the opportunity cost—which the accountant would not include in total costs—is $38,000 ($35,000 plus $3,000). To get economic profits, we have to subtract this opportunity cost from accounting profits. Thus, economic profits would be only $2,000.

Economic profits are used by economists because they measure the incentive the owner of the firm has to stay in business versus doing something else. In this case, with $2,000 in economic profits, the owner has an incentive to stay in the business. But if the owner could earn $39,000 managing a video rental store, then economic profits for the bakery would be −$2,000 (40,000 − 39,000 − 3,000), and the owner would have an incentive to run the video store. Even though accounting profits at the bakery were $40,000, the owner would have an incentive to go to work elsewhere because economic profits would be −$2,000. Thus, economic profits are a better measure of incentives than accounting profits, and this is why economists focus on economic profits. When we refer to profits in this book, we mean economic profits because we are interested in the incentives firms have to either enter or exit an industry.

Observe that if the bakery owner could earn exactly $37,000 at the video rental store, then economic profits at the bakery would be zero. Then the owner would be indifferent on economic considerations alone between staying in the bakery business or going to work for the video rental company. The term **normal profits** refers to the amount of accounting profits that exist when economic profits are equal to zero. In this last case, normal profits would be $40,000.

normal profits: the amount of accounting profits when economic profits are equal to zero.

The Equilibrium Number of Firms. The long-run equilibrium model predicts that there will be a certain number of firms in the industry. The equilibrium number of firms will be such that there is no incentive for more firms to enter the industry or for others to leave. But how many firms is this? If the minimum point on the average cost curve of the typical firm represents production at a very small scale, then there will be many firms. That is, many firms will each produce a very small amount. If the minimum point represents production at a large scale, then there will be fewer firms; that is, a few firms will each produce a large amount.

To see this, consider the hypothetical case where all firms are identical. For example, if the minimum point on the average total cost curve for each firm in the grape industry occurs at 10,000 tons and the size of the whole market is 100,000 tons, then the model predicts 10 firms in the industry. If the quantity where average total cost is at a minimum is 1,000 tons, then there will be 100 firms. If in the latter case the demand for grapes increases and brings about a new long-run equilibrium of 130,000 tons, then the number of firms in the industry will rise from 100 to 130.

■ **Entry Combined with Individual Firm Expansion.** Thus far, we have described the growth of an industry in terms of the increase in the number of firms. In the short run, immediately after a change in demand, there is no entry or exit; then entry takes place and the industry moves toward a new equilibrium in the long run. Recall from Chapter 8 that something else can occur in the long run but not in the short run: In the short run, a firm cannot expand its size by investing in new capital, but in the long run it can expand.

In reality, industries usually grow by a combination of the expansion of existing firms and the entry of new firms. For example, this was what happened in the expedited package express industry, which grew both because UPS and other firms entered and because Federal Express expanded.

The expansion of an existing firm can occur under one of two conditions. First, the original size of the firm may be smaller than the minimum efficient scale, so the firm may be able to lower its average costs while producing more units. Second, a change in technology or in the prices of inputs may change the cost function of the firm, pushing the minimum efficient scale to a larger number of units. Note that if the firm is already producing at the minimum long-run average total cost, then an increase in demand will not affect the size of the firm, and you will observe only entry of new firms into the industry.

A Decrease in Demand

The long-run competitive equilibrium model can also be used to explain the decline of an industry. Suppose there is a shift in the demand curve from D to D', as illustrated in the top right graph in Figure 9.5. This causes the market price to fall. The lower market price (P') causes existing firms to cut back on production in the short run: As production decreases, marginal cost falls until it equals the new lower price for each firm. However, the firms are now running losses. As shown in the top left graph of Figure 9.5, profits drop below zero.

With profits less than zero, firms now have an incentive to leave the industry. As they leave, the market supply curve shifts to the left from S to S' as shown in the bottom right graph of Figure 9.5. This causes the price to rise again. The end of the process is a new long-run equilibrium, as shown in the bottom left graph of Figure 9.5. In the long run, fewer firms are in the industry, total production in the industry is lower, and profits are back to zero.

Shifts in Cost Curves

Our analysis of the rise and fall of an industry thus far has centered around shifts in demand. But new technologies and ideas for new products that reduce costs can also cause an industry to change. The long-run competitive equilibrium model can also be used to explain these changes, as shown in Figure 9.6.

The case of cost-reducing technologies—as when Wal-Mart introduced checkout counter scanners—can be handled by shifting down the average total cost curve along with the marginal cost curve, as shown in Figure 9.6. This will lead to a situation of positive profits because average total cost falls below the original market price P. If other firms already in the industry adopt similar cost-cutting strategies, the market price will fall to P', but profits will still be positive, as shown in Figure 9.6. With positive profits, other firms will have incentives to enter the industry with similar

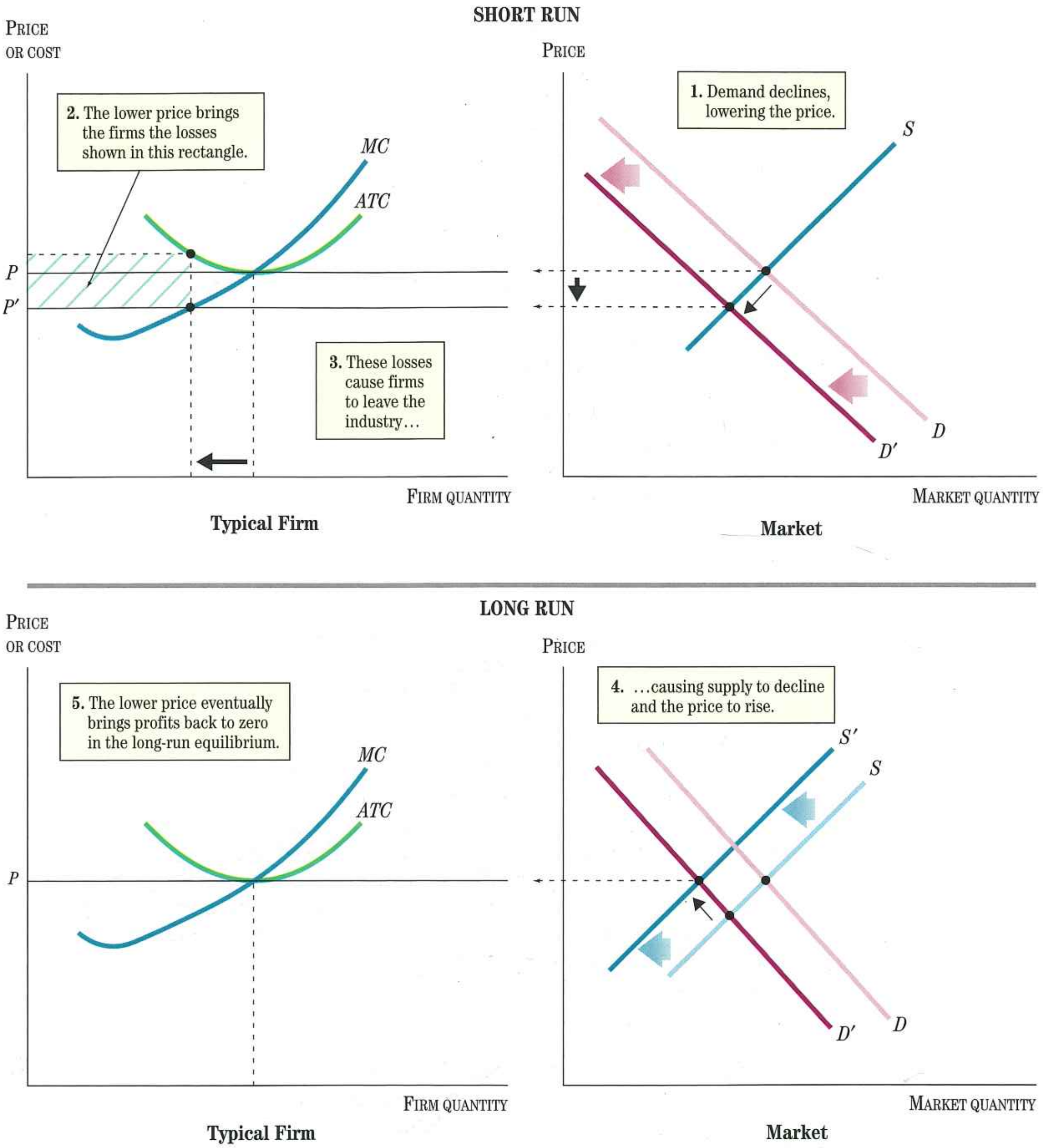

SHORT RUN
PRICE OR COST
2. The lower price brings the firms the losses shown in this rectangle.
MC
ATC
P
P′
3. These losses cause firms to leave the industry…
FIRM QUANTITY
Typical Firm
PRICE
1. Demand declines, lowering the price.
S
D
D′
MARKET QUANTITY
Market
LONG RUN
PRICE OR COST
5. The lower price eventually brings profits back to zero in the long-run equilibrium.
MC
ATC
P
FIRM QUANTITY
Typical Firm
PRICE
4. …causing supply to decline and the price to rise.
S′
S
D′
D
MARKET QUANTITY
Market

Figure 9.5
The Decline of an Industry after a Shift in Demand
In the short run, a reduction in demand lowers the price from *P* to *P′* and causes losses. Firms leave the industry, causing prices to rise back to *P*. In the long run, profits return to zero, the number of firms in the industry has declined, and the total quantity produced in the industry has fallen.

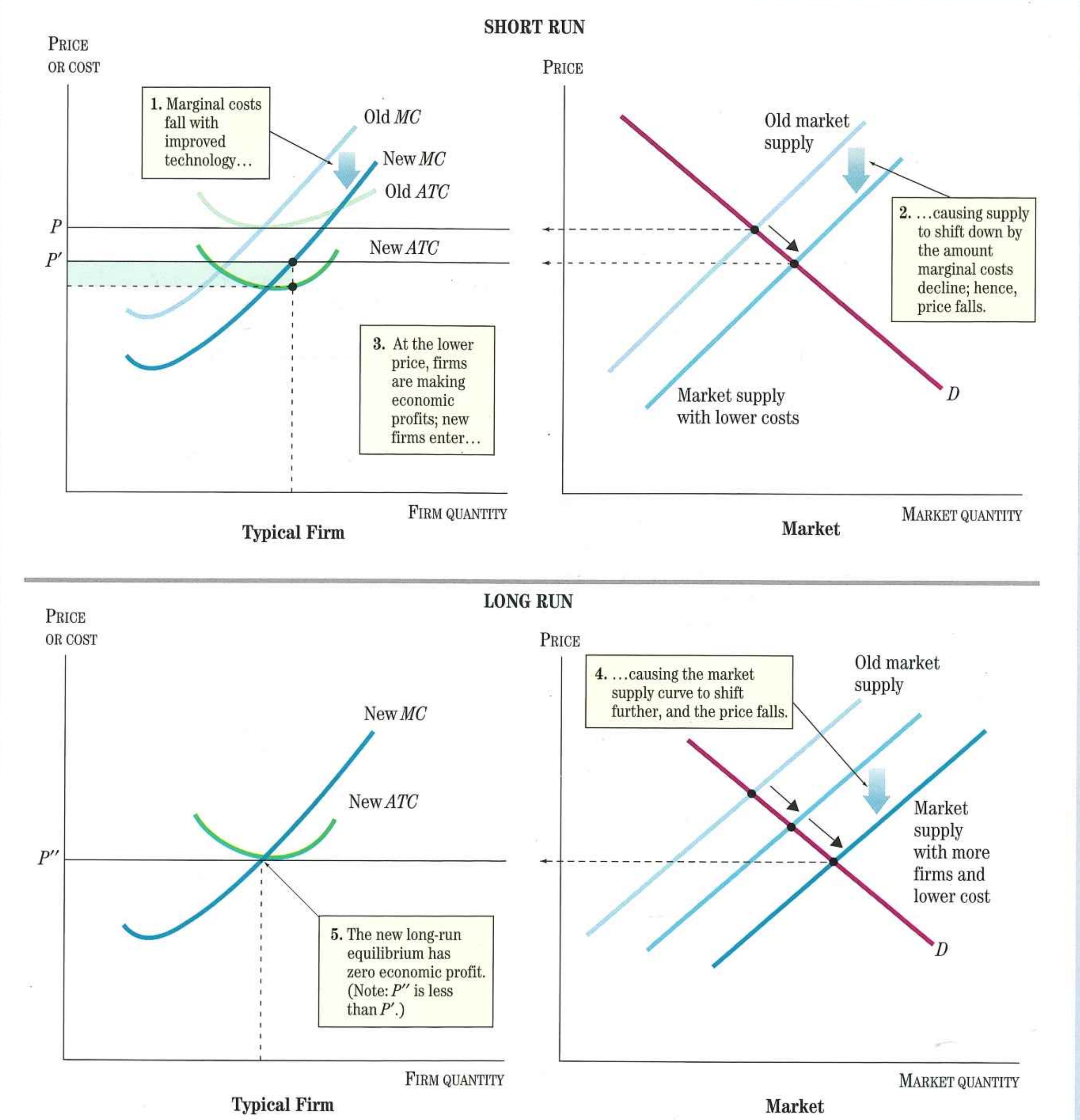


Figure 9.6
Effect of a Reduction in Costs
A new technology reduces costs and shifts the typical firm's *ATC* and *MC* curves down. The market supply curve shifts down by the same amount as the shift in marginal cost if other firms in the industry adopt the new technology right away. But because there are economic profits, new firms have incentives to enter the industry. As shown in the lower left graph, in the long run, profits return to zero.

Economics in Action

Digital Cameras and the Future of Silver Halide Film

The long-run competitive equilibrium model is useful not only for explaining what happened to an industry in the past, but also for predicting what will happen to an industry in the future. Such predictions can help guide investment or career decisions.

The photographic film industry provides an interesting example of an industry undergoing change based on other industry changes. Photographic film, called silver halide film, was developed over 100 years ago and can give extremely high resolution and detail. The same technology is used for x-ray pictures, where detailed views are needed to detect hairline fractures and the like. This amazing film technology has brought enjoyment and better health to many millions of people.

But in recent years, digital cameras have begun to outsell cameras that use film. With a digital camera, you can take snapshots and load the images directly into a desktop computer. You can enlarge the images yourself, or have fun coloring your hair purple and e-mailing the image to your mother or father. Digital cameras are different from analog cameras in a very important way: *Digital cameras do not use film.* The more people use digital cameras, the less film they will buy. Thus, we can predict that the demand for photographic film and developing services will decline in the future as digital cameras improve and become cheaper. The demand curve will shift to the left, profits will decline, and firms will exit the industry. Figure 9.5 tells the story.

But is that the whole story? Film sales and film developing services may be declining rapidly, but many people still want to print their favorite images. And while some of the prints will be made on home inkjet printers, there is still a large demand for printing from retailers due to the high cost and amount of time it takes to make a print at home. These prints, whether they come from film or a digital image, are produced on silver halide–based photographic paper. Not only is the demand for photographic paper and printing services likely to grow, but the Silver Institute estimates that silver usage for these products will rise from 46.0 million troy ounces in 2000 to 60.1 million troy ounces in 2008.[1]

Naturally, other things could change. If the cost of printing high-quality photographs at home declines, fewer people might go outside to have their photos printed. And who knows what shifts another new entry into this market—the camera phone—might bring? Regardless of what happens, the economic model can help us determine the impact of changing technology and demand on industries in the economy.

[1] Don Franz, "The Global Silver Halide Photographic Market," *Silver News*, First Quarter 2004, www.silverinstitute.org.

FUJIFILM
OFFICIAL IMAGING SPONSOR OF WORLD CUP FRANCE 98
New
FUJIX Digital Camera
DS-300
DS-300 has five exposure modes.
the DS-300 delivers
a 6x telephoto setting
the DS-300 records images
THE CREATIVITY AND FUN ARE BUILT IN

If you've ever wondered whether a digital camera can combine quality and simplicity, the DS-300 has all the right answers. In a moment, you'll discover that it couldn't be easier to use. And a glance at the image quality will tell you that the performance and resolution are comparable to those of our most advanced digital SLRs.

Will digital cameras continue to shift the demand curve for film?

technologies. As the market supply curve shifts out after more firms enter the industry, the price falls further to P'', and eventually competition brings economic profits back to zero.

If new entrants drive economic profits to zero in the long run, then what incentives do firms have to develop cost-cutting technologies? The answer is that the economic profits in the short run can be substantial. Wal-Mart may have made hundreds of millions of dollars in economic profits before the competition eroded them. Hence, Wal-Mart benefited for a while from cost-cutting innovations. No idea will generate economic profits forever in a competitive market, but the short-run profits can still provide plenty of incentive.

REVIEW

- Entry and exit of firms in search of profit opportunities play a key role in the long-run competitive equilibrium model. The decision to enter or exit an industry is determined by profit potential. Positive economic profits will attract new firms. Negative economic profits will cause firms to exit the industry. In long-run equilibrium, economic profits are zero.
- The market supply curve shifts as firms enter or exit the industry. With more firms, the market supply curve shifts to the right. With fewer firms, the market supply curve shifts to the left.
- The model can be used to explain the rise and fall of many different industries, whether due to shifts in demand or to shifts in cost curves.

CASE STUDY

How Does the Model Explain the Facts?

For the purposes of using the long-run competitive equilibrium model in practice, economists usually need to narrow their focus.

Consider, for example, changes in the grape industry. The grape industry includes many different types of products. Some grape types are used to make raisins, other grapes are grown to be table grapes, and many different types of grapes are used for wine. Even within each of these categories, there are particular styles of grapes.

A good case study is the Zinfandel grape. This is one grape that is still largely confined to the United States, so we do not need to consider developments throughout the world. It is used to produce a particular type of wine, also called Zinfandel. Figure 9.7 shows the rise and fall in the price of Zinfandel grapes. The price more than tripled between 1985 and 1988. Then, nearly as sharply, the price declined from 1989 to 1991.

Figure 9.8 shows what happened to production in the Zinfandel grape industry during this period. The industry grew rapidly from 1985 to 1988. New vineyard acreage entering the industry each year more than tripled. In 1988, industry growth slowed and the number of acres of new vineyards declined sharply. By 1991, growth was close to zero, with only a handful of new vineyards entering.

How can we explain this huge rise and subsequent slowdown of the Zinfandel grape industry? The most likely explanation centers around the discovery of a new product. In the mid-1980s, it was discovered that Zinfandel grapes could be used to produce a new type of wine called "white Zinfandel," which proved to be very popular. Previously the grape had been used only to produce a heavy red wine that was less popular. This discovery greatly increased the demand for the Zinfandel grape.

According to the competitive equilibrium model, the increase in demand would be represented as a shift of the demand curve to the right, exactly as shown in Figure

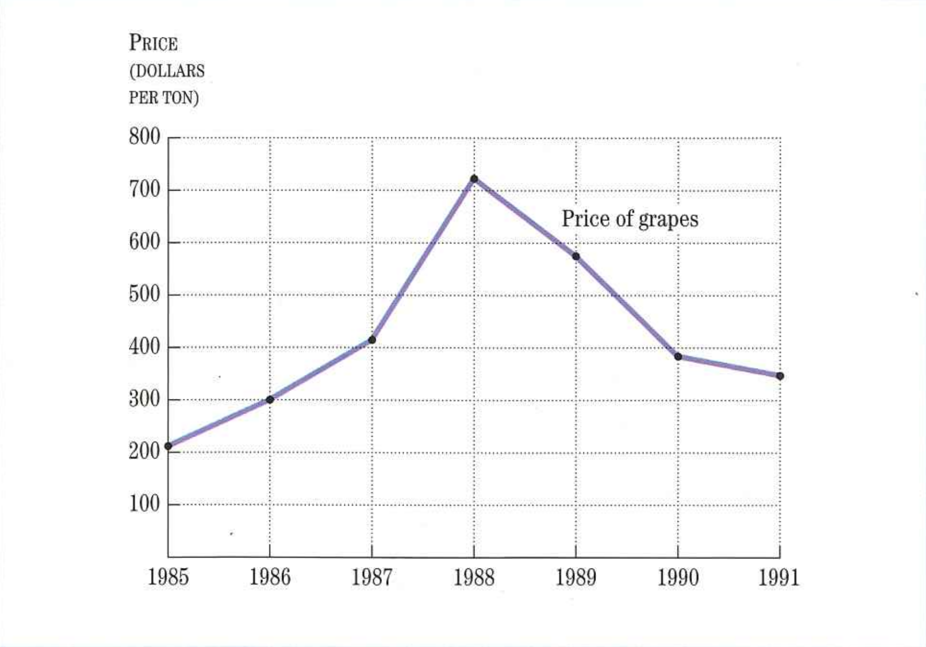


Figure 9.7
The Price of Grapes, 1985–1991
The price of Zinfandel grapes rose in the late 1980s because tastes shifted in favor of white wine made from these grapes. The higher price raised profits at Zinfandel vineyards and the number of vineyard acres increased as a result, as shown in Figure 9.8.

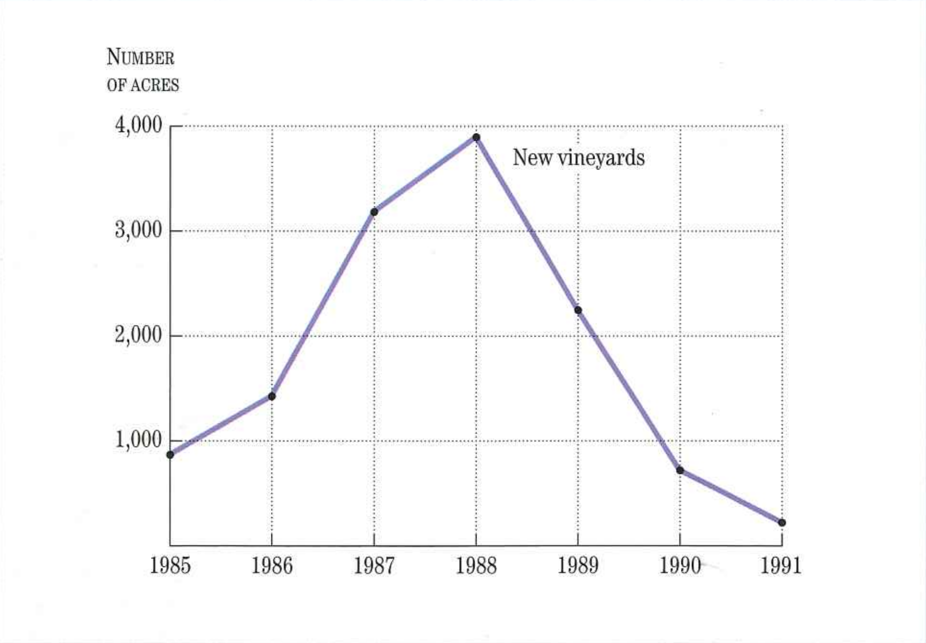


Figure 9.8
New Vineyards in the Grape Market, 1985–1991
As profits from growing Zinfandel grapes rose, the number of vineyards increased. The resulting increase in supply eventually lowered the price, which lowered profits, and the number of new vineyards declined. It takes several years for a new vineyard to start producing grapes.

9.4. The model would predict a rise in the price of grapes, and that is exactly what happened. The price rose from 1985 to 1988 after the demand curve shifted, as shown in Figure 9.7. With the higher price, the profits from Zinfandel grape production would increase; thus, the model predicts the entry of new firms. In fact, as shown in Figure 9.8, Zinfandel vineyard acreage did increase sharply in this same time period.

READING THE NEWS ABOUT... Cycles in the Grape Industry

The following article discusses fluctuations in the California grape industry. It shows how a positive demand shock, in this case resulting from the discovery of potential health benefits of drinking wine, leads to improved profits for domestic wine producers, in turn leading more firms to enter into production. Over time, the increase in supply drove prices down and moved the wine industry toward long-run equilibrium. Recently, however, increases in wine imports from countries like South Africa, Chile, and Australia, as well as competition from other types of alcoholic beverages, have reduced the demand for domestic wine, causing prices to drop as the industry again moves toward long-run competitive equilibrium.

"What We Have Is Insufficient Demand for Wine Grape Supplies"

westernfarmpress.com
Feb. 4, 2003 12:00 PM, Harry Cline

California's wine grape industry is not going to heck in an oak barrel or a stainless steel tank.

"What we have is insufficient demand for the existing supplies," pronounced wine industry expert Barry Bedwell in explaining the current economic plight of California wine grapes in his best impersonation of actor Strother Martin. As you recall, Martin was immortalized for his famous line, "What we have here is a failure to communicate" from the movie Cool Hand Luke.

Bedwell, California wine broker coordinator for Joseph W. Ciatti Co. and Jon Fredrikson of Gomberg, Fredrikson and Associates avoided the G (glut) word like a glass of bad White Grenache when they reported on the status of the industry to a rapt crowd of more than 1,000 at the 9th annual United Wine and Grape Symposium recently in Sacramento, Calif.

Fredrikson and Bedwell are two of the most respected wine industry analysts. Bedwell's opening comment about "insufficient demand" was calculated to net a laugh from the audience, and it did.

According to the model, in the long run, the increased supply from the new entrants to the industry should lower the price. As shown in Figure 9.7, the price did fall after 1988.

Again according to the model, the lower price should reduce profit opportunities, and the number of new entrants should decline. Sure enough, the data show that the number of new entrants peaked in 1988 and then declined.

At the end of this process, the price had returned to near the original price and the number of new entrants was close to zero. Apparently a new equilibrium was reached. Overall, the facts during this episode of change in the industry seem to be explained quite well by the model.

REVIEW

- To apply the long-run competitive equilibrium model, economists focus on a single industry with a clear shift in demand or supply.
- Case studies like that of the Zinfandel grape industry show that the model works.

However, it has not been a laughing matter for most wine grape growers and many wineries over at least the past two years as acreage coming into production has soared, creating an oversupply in many varietals and sending prices plummeting; imports taking an increasingly bigger share of the U.S. wine market and wine finding tough going in getting shelf space from beers and spirits.

No Wine Glut

However, there is no California wine glut, Bedwell said. There are oversupplies of some varietals like Cabernet Sauvignon, Pinot Noir and perhaps Syrah. However, supply and demand are getting closer to balances for Chardonnay, red Zinfandel, Sauvignon Blanc, Merlot and White Zinfandel.

The California wine industry is in a "down cycle," not in wine glut, according to Bedwell, after experiencing a decade of phenomenal growth following the 1991 French Paradox. That broadcast heralded the health benefits of moderate wine drinking. Sales have increased by 75 million cases since then.

"What I think Jon and I were trying to do in our industry assessments is counter the overwhelming negative publicity of the wine industry over the past few months that is ignoring the cyclical aspects of this industry," said Bedwell.

"This is a remarkably strong, $1.5 billion a year industry," Bedwell said.

"Is there an oversupply of wine grapes? Absolutely," said Bedwell, who said the value of wine grapes dropped $200 million last year and likely will drop another 7 percent to 10 percent this season as the industry works off oversupplies.

"The California situation is not a glut," echoed Fredrikson. "A glut is the wine lake in Europe where growers are producing wine that is not intended for sale as wine but to be used as gasohol as part of a social program."

Bedwell said the California wine industry has become notorious for overreacting. When things are good, new plantings quickly catch up with grape and wine demand and an oversupply situation is created, even with growing wine sales. When things go bad, growers become too aggressive with bulldozers in taking out vineyards.

- An increase in demand caused the price of Zinfandel grapes to rise. Grape-producing firms saw a profit opportunity and entered the industry, transforming existing land into vineyards. The supply of grapes began to increase. As a result, the price started to come back down again.

Minimum Costs per Unit and the Efficient Allocation of Capital

If firms can enter or exit a competitive industry, as assumed in this model, then there are several other attractive features of the competitive market that we can add to those discussed in Chapter 7.

Average Total Cost Is Minimized

In the long-run equilibrium, average total cost is as low as technology will permit. You can see this in Figures 9.3, 9.4, 9.5, and 9.6. In each case, the typical firm produces a quantity at which average total cost is at the *minimum point* of the firm's average total cost curve. This amount of production must occur in the long-run equilibrium because profits are zero. For profits to be zero, price must equal average total cost ($P = ATC$). The only place where $P = MC$ and $P = ATC$ is at the lowest point on the *ATC* curve. At this point, costs per unit are at a minimum.

In the long run, firms can expand or contract as well as enter or exit an industry. As they expand or contract, their costs are described by the long-run *ATC* curve. Thus, in the long-run competitive equilibrium, firms operate at the lowest point of the long-run average total cost curve.

That average total cost is at a minimum is an attractive feature of a competitive market where firms are free to enter and exit. It means that goods are produced at the lowest cost, with the price consumers pay equal to that lowest cost. If firms could not enter and exit, this attractive feature would be lost.

Efficient Allocation of Capital among Industries

An efficient allocation of capital among industries is also achieved by entry and exit in competitive markets. Entry of firms into the Zinfandel grape industry, for example, means that more capital has gone into that industry, where it can better satisfy consumer tastes, and less capital has gone into some other industry.

In the case of a declining industry, capital moves out of the industry to other industries, where it is more efficiently used. For example, capital moved away from the beef industry toward the chicken industry when the former contracted and the latter expanded in recent years. Thus, the long-run competitive equilibrium has another attractive property: Capital is allocated to its most efficient use. Again, this property is due to the free entry and exit of firms. If entry and exit were limited or if the market were not competitive for some other reason, this advantage would be lost.

REVIEW

- In a long-run competitive equilibrium, firms operate at the minimum point on their long-run average total cost curves and capital is allocated efficiently across different industries.
- Minimum-cost production is a benefit to society of the competitive market with free entry and exit.

External Economies and Diseconomies of Scale

In Chapter 8 we introduced the concept of economies and diseconomies of scale for a firm. A firm whose long-run average total cost declines as the firm expands has economies of scale. If long-run average total cost rises as the firm expands, there are diseconomies of scale. Economies and diseconomies of scale may exist for whole industries as well as for firms, as we now show.

External Diseconomies of Scale

When the number of firms in the Zinfandel grape industry increases, the demand for water for irrigation in grape-growing regions also increases, and this may raise the price of water in these regions. If it does, then the cost of producing grapes increases. With the marginal cost of each grape producer increasing, the supply curve for each firm and for the industry or the market shifts up and to the left. Even though no single firm's decision affects the price of water for irrigation, the expansion of the industry does.

This is shown in the market supply and demand curves in Figure 9.9. Suppose there is a shift in the demand curve from D_1 to D_2. As the industry expands, more firms enter the industry and the supply curve shifts to the right from S_1 to S_2. Because the marginal cost at each firm rises as the industry expands, the supply curve does not shift to the right by as much as the demand curve shifts. Thus, the intersection of the demand curve D_2 and the supply curve S_2 occurs at a higher price than the intersection of S_1 and D_1.

We could consider a further shift in demand to D_3, leading to a shift in supply to S_3. This would result in yet another long-run equilibrium at a higher price because average total cost is higher. Observe that as successive market demand curves intersect successive market supply curves, the price rises and quantity rises; an upward-sloping **long-run industry supply curve** is traced out. We call the phenomenon of an upward-sloping long-run industry supply curve **external diseconomies of scale.** The word *external* indicates that cost increases are external to the firm, due, for example, to a higher price for inputs (such as water) to production. In contrast, the diseconomies of scale considered in Chapter 8 were internal to the firm, due, for example, to increased costs of managing a larger firm; they can be called *internal diseconomies of scale* to distinguish them from the external case.

long-run industry supply curve: a curve traced out by the intersections of demand curves shifting to the right and the corresponding short-run supply curves.

external diseconomies of scale: a situation in which growth in an industry causes average total cost for the individual firm to rise because of some factor external to the firm; it corresponds to an upward-sloping long-run industry supply curve.

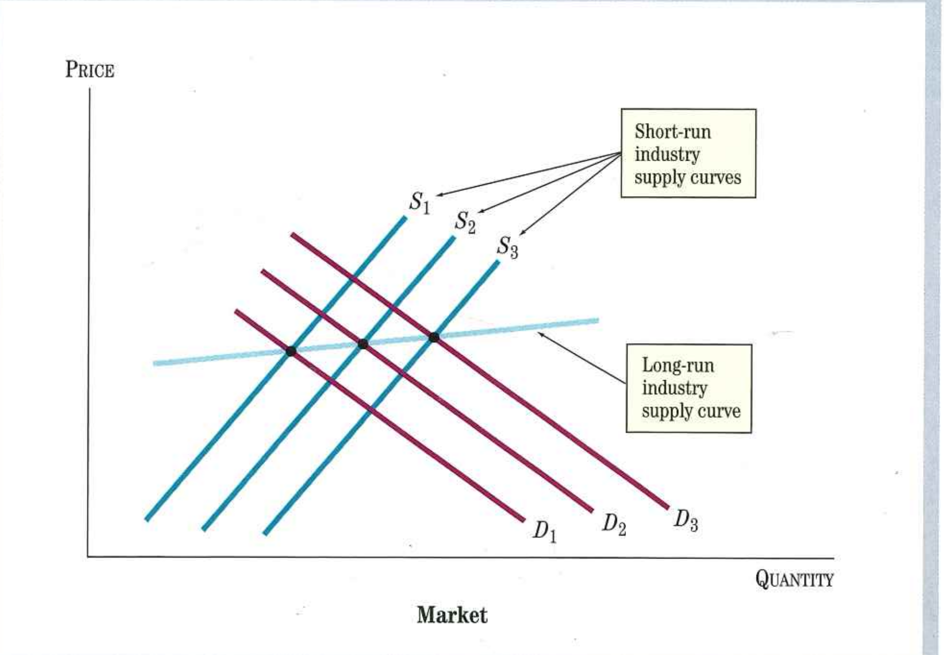


Figure 9.9
External Diseconomies of Scale
As demand increases and more firms enter the industry, each firm's costs increase, perhaps because the prices of inputs to production rise. The higher costs tend to limit the shift of the market supply curve to the right when new firms enter. The long-run industry supply curve slopes up, a phenomenon that is called external diseconomies of scale.

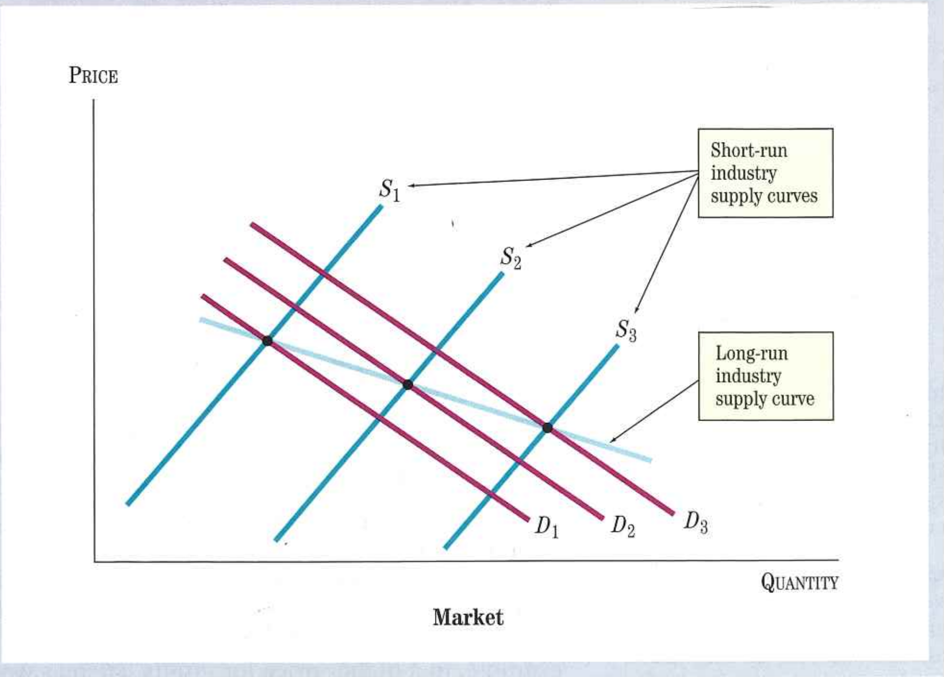


Figure 9.10
External Economies of Scale
As demand expands and more firms enter the industry, each firm's costs decline, which causes the supply curve to shift to the right by even more than it would as a result of the increase in the number of firms. The long-run industry supply curve is thus downward-sloping, a phenomenon that is called external economies of scale.

External Economies of Scale

external economies of scale: a situation in which growth in an industry causes average total cost for the individual firm to fall because of some factor external to the firm; it corresponds to a downward-sloping long-run industry supply curve.

External economies of scale are also possible. For example, the expansion of the Zinfandel grape industry might make it worthwhile for students at agricultural schools to become specialists in Zinfandel grapes. With a smaller industry, such specialization would not have been worthwhile. The expertise that comes from that specialization could reduce the cost of grape production by more than the cost of hiring the specialist. Then as the industry expands, both the average total cost and the marginal cost for individual firms may decline.

The case of external economies of scale is shown in Figure 9.10. Again, suppose there is a shift in the demand curve from D_1 to D_2. When the industry expands, the market supply curve shifts out from S_1 to S_2, or by *more* than the increase in demand, so that the price falls. The reason the market supply curve shifts more than the market demand curve is that marginal cost at each firm has declined as the number of firms in the industry has increased. This larger shift in supply compared to demand is shown in Figure 9.10. Thus, the price falls as the industry expands.

With additional shifts in demand from D_2 to D_3, the market demand curves intersect with successive market supply curves at lower prices, resulting in a long-run industry supply curve that is downward-sloping. Again, the word *external* is used to distinguish these economies that occur outside the firm from those that are internal to the firm.

Our example of the reason why external economies occur is that a larger industry allows for opportunities for *specialization*—grape-growing specialists who then provide services to the industry. There are other such examples. The expansion of the personal computer industry made it worthwhile for many small specialized firms servicing personal computer manufacturers to emerge. With a smaller-scale industry, this would not have been possible.

Internal economies or diseconomies of scale occur when a *single firm* expands (Chapter 8). *External* economies or diseconomies of scale occur when an *industry* expands (Chapter 9).

Note the difference between internal and external economies of scale. The expansion of a single firm can generate internal economies of scale with the number

External Economies of Scale
As an industry expands in size, firms in other industries have incentives to develop new products to service the industry. These new products reduce average total cost in the industry, thereby giving rise to economies of scale, as illustrated by the development of special electronic scanners for use by the expanding express delivery service industry (left). The new ideas may in turn be used to reduce costs in other industries, as illustrated by the use of electronic scanners for self-service checkout in the retail food industry (right).

of firms in the industry fixed because individuals within the firm can specialize. The expansion of an industry can generate external economies of scale even if the size of each firm in the industry does not increase. As an industry expands, firms might even split up into several specialized firms, each concentrating on one part of the specialized work.

The Standard Assumption: A Flat Long-Run Industry Supply Curve

In the examples used in the previous sections of this chapter, there were neither external economies of scale nor external diseconomies of scale. To convince yourself of this, look back to the graphs in the lower right-hand panels of Figures 9.4 and 9.5. You will see that the market price in the long-run equilibrium after the shift in demand is the same as before the shift in demand. If you draw a line between the intersection points of the supply and demand curves in those graphs, you will get a flat line. Thus, the long-run industry curve will be perfectly horizontal, in contrast to the upward slope in Figure 9.9 and the downward slope in Figure 9.10. The intersections of the shifting demand and supply curves trace out neither an upward-sloping long-run industry supply curve—the case of external diseconomies of scale—nor a downward-sloping long-run industry supply curve—the case of external economies of scale. The assumption of a flat long-run industry supply curve is the standard one economists use to study industries where neither type of external scale effect is known to occur.

External and Internal Economies of Scale Together

In practice, it is possible for external and internal economies of scale to occur at the same time in one industry. When an industry grows in scale through the addition of new firms, it is common for the typical firm in the industry to expand its scale.

Federal Express has grown in size at the same time that more firms have entered the industry. Through its larger size, Federal Express has achieved internal economies of scale (for example, by spreading the costs of its computer tracking system over more deliveries), and the larger industry as a whole has benefited from external economies of scale (as illustrated by the scanners shown in the photo).

REVIEW

- External diseconomies of scale occur when an expansion of an industry raises costs at individual firms, perhaps because of a rise in input prices.
- External economies of scale arise when expansion of an industry lowers costs at individual firms in the industry. Opportunities for specialization for individuals and firms serving the industry are one reason for external economies of scale.

Conclusion

In this chapter, we have addressed one of the most pervasive and perplexing realities of a market economy: the changes that occur when whole industries rise or fall over time. As consumer tastes change and new ideas are discovered, such changes are an ever-present phenomenon in modern economies around the world.

The model we have developed in this chapter to explain such changes extends the competitive equilibrium model we developed in Chapters 5, 6, 7, and 8 to allow for the entry or exit of firms into or out of an industry. Because such entry or exit usually takes time, we emphasize that this modification applies to the long run. Profits motivate firms to enter or exit an industry. Profits draw firms into the industry over time, whereas losses cause firms to leave. As firms enter, the industry expands. As firms leave, the industry declines. In the long-run equilibrium, profit opportunities have disappeared, and entry or exit stops.

In Chapters 10 and 11, we begin to leave the realm of the competitive market. We will develop models of the behavior of monopolies and other firms for which the assumption of a competitive market is not accurate. In the process, we will see that many of the results we have obtained with competitive markets in this chapter are no longer true.

However, many of the ideas and concepts developed in this and the previous few chapters on the competitive model will be used in these chapters. The cost curve diagram will reappear in the model of monopoly in Chapter 10; the idea of entry and exit will reappear in Chapter 11.

As we consider these new models and new results, we will use the models of this chapter as a basis of comparison. A central question will be: "How different are the results from those of the long-run competitive equilibrium model?" Keep that question in mind as you proceed to the following chapters.

KEY POINTS

1. Economic history is filled with stories about the rise and fall of industries. Industries grow rapidly when cost-reducing technologies are discovered or demand increases. They decline when demand decreases.
2. Because of reduced transportation and communication costs, most industries today are global.
3. The economists' competitive equilibrium model assumes that firms are price-takers.

4. The long-run competitive equilibrium model also assumes that firms enter or exit an industry until economic profits are driven to zero.
5. The long-run competitive equilibrium model can be used to explain many facts about the rise and fall of industries over time.
6. In the long run, the competitive equilibrium model implies that after entry and exit have taken place, average total costs are minimized and capital is allocated efficiently among industries.
7. Industries may exhibit either external economies of scale, when the long-run industry supply curve slopes down, or external diseconomies of scale, when the long-run industry supply curve slopes up.

KEY TERMS

industry
long-run competitive equilibrium model
free entry and exit
long-run equilibrium
accounting profits
economic profits
normal profits
long-run industry supply curve
external diseconomies of scale
external economies of scale

QUESTIONS FOR REVIEW

1. What are three possible sources of the rise of industries? Of the fall of industries?
2. What is the difference between economic profits and accounting profits?
3. Why do firms enter an industry? Why do they exit?
4. Why does the market supply curve shift to the right when there are positive profits in an industry?
5. Why is "zero profits" a condition of long-run equilibrium?
6. What does the demand curve look like to a single firm in a competitive industry?
7. Why are average total costs minimized in a long-run competitive equilibrium?
8. What are external economies of scale? How do they differ from internal economies of scale?
9. How does the long-run/short-run distinction differ when applied to a firm versus an industry?

PROBLEMS

1. Suppose corn farmers in the United States can be represented by a competitive industry with no economies or diseconomies of scale. Describe how this industry would adjust to an increase in demand for corn. Explain your answer graphically, showing the cost curves for the typical farmer as well as the market supply and demand curves. Distinguish between the short run and the long run.
2. Suppose the government gives a subsidy to textile firms, paying each firm a specific amount per unit of production. What will happen to output and the number of firms in the short run and the long run?
3. Sketch a diagram showing the costs and price of the typical price-taking firm in long-run equilibrium. Suppose a technology is invented that reduces average total cost and marginal cost. Draw this new situation. Describe how the industry adjusts. How will the long-run equilibrium price change? What happens to the number of firms in the industry?
4. Suppose the government imposes a sales tax on a good sold by firms in a competitive industry. Describe what happens to the price of the good in the short run and in the long run when firms are free to enter and exit. What happens to the number of firms in the industry and to total production in the industry?
5. Consider a typical carpet-cleaning firm that currently faces $24 in fixed costs and an $8 hourly wage for workers. The price it gets for each office cleaned in a large office building is $48 at the present long-run equilibrium. The production function of the firm is shown in the following table:

Number of Offices Cleaned	Hours of Work
0	0
1	5
2	9
3	15
4	22
5	30

   a. Find marginal costs and average total costs for the typical firm.
   b. How many offices are cleaned by the typical firm in long-run equilibrium?

c. Suppose there is an increase in demand. Describe the process that leads to a new long-run equilibrium. What is the new price in the long-run equilibrium? What is the quantity produced by the typical firm? Draw the market demand and supply curves before and after the shift. (Assume that the hourly wage remains at $8 per hour.)

d. Now assume that the increased number of cleaning firms causes a rise in the hourly wage from $8 to $9. How would your answer differ from that for part (c)? In particular, compare the equilibrium price and the market demand and supply curves.

6. Compare and contrast economic profits, accounting profits, and normal profits.

7. Many young children drink from sippy cups—plastic cups with a spout that prevents the liquid contents from spilling. Recently pediatric dentists have attacked the use of sippy cups, saying that they may contribute to the formation of cavities.

   a. Graphically show the competitive market for sippy cups before and after this news is released to the public. What would you predict will happen to equilibrium price and quantity in the short run?

   b. Now graphically show the long-term equilibrium of the sippy cup market. Which curves shift? Why? Compare the long-term equilibrium price and quantity with its original counterparts.

8. What is the difference between the short run and the long run for a firm (as described in Chapter 8) and the short run and the long run for an industry (as described in this chapter)?

9. What is the incentive for one firm in a competitive industry to pursue cost-cutting measures? What will happen in the long run?

10. This problem combines changes in capital at each firm over the long run with entry and exit of firms into or out of an industry in the long run. Given the data in the table for a typical firm in a competitive industry (with identical firms), sketch the two short-run average total cost curves (ATC_1 and ATC_2) and the two marginal cost curves (MC_1 and MC_2).

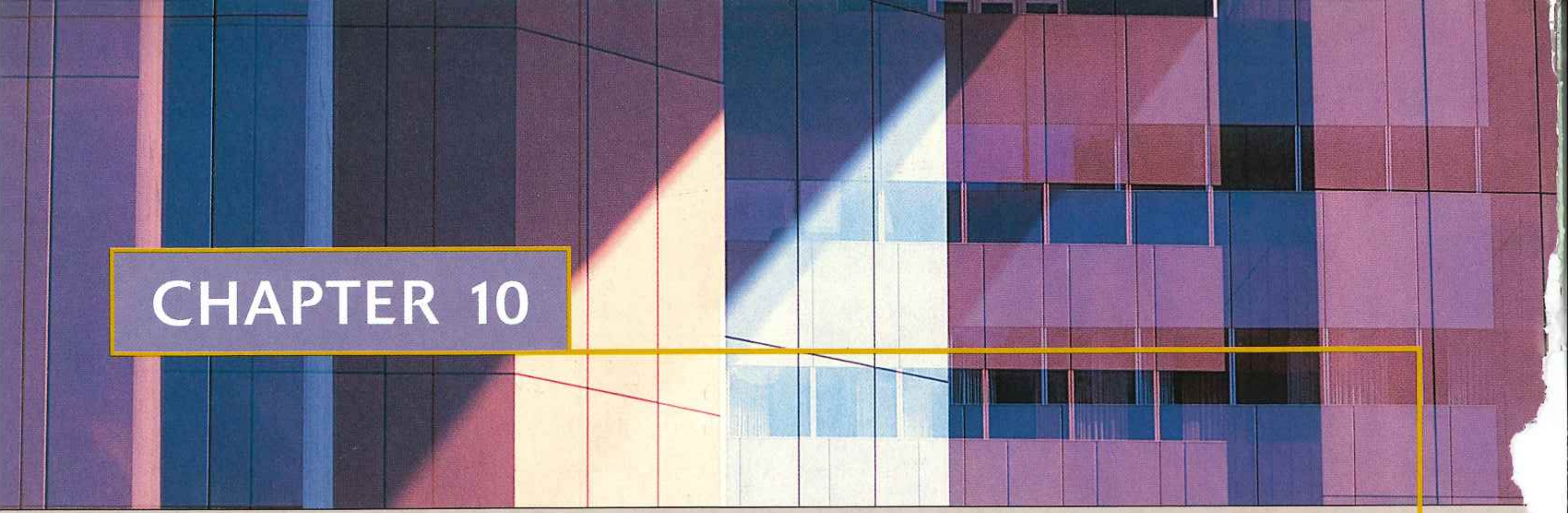
    a. Suppose the price is $9 per unit. How much will the firm produce, and with what level of capital?

    b. Suppose the firm is currently producing with 2 units of capital. If the price falls to $7 per unit, will the firm contract when it is able to change its capital? Why?

    c. What is the long-run industry equilibrium price and quantity for the typical firm? If there is a market demand of 4,000 units at that price, how many of these identical firms will there be in the industry?

    d. Why might the firm operate with 2 units of capital in the short run if the long-run equilibrium implies 1 unit of capital?

	Costs with 1 Unit of Capital		**Costs with 2 Units of Capital**	
Quantity	ATC_1	MC_1	ATC_2	MC_2
1	7.0	5	10.0	6
2	5.5	4	7.5	5
3	4.7	3	6.3	4
4	4.5	4	5.5	3
5	4.6	5	5.0	3
6	4.8	6	4.8	4
7	5.1	7	4.9	5
8	5.5	8	5.0	6
9	5.9	9	5.2	7
10	6.3	10	5.5	8
11	6.7	11	5.8	9
12	7.2	12	6.2	10



CHAPTER 10

Monopoly

Sending the board game Monopoly to prospective customers at the start of the 1990s was how Advanced Micro Devices called attention to Intel's monopoly. Both Advanced Micro Devices and Intel produce the central processor—the brains—of personal computers. Intel had a monopoly because it was the sole producer of the 386 computer chip, the most popular central processor for personal computers in the early 1990s. Intel's researchers had invented the chip. The company created the monopoly legally by obtaining patents and copyrights from the government that gave it the sole right to produce the chip. No other firm could compete with Intel unless Intel gave permission, and Intel did not want to give permission to Advanced Micro Devices. As frequently happens, however, Advanced Micro Devices designed its own 386 chip, a clone of Intel's that did not infringe on Intel's patents. By 1991, Advanced Micro Devices was ready to compete with Intel and used the Monopoly game to help launch its product. But by the time it had done so, Intel had invented a more advanced chip—the 486—and created yet another monopoly. Nevertheless, Advanced Micro Devices soon got into the new act, developing chips that virtually replicated Intel's 486. Today the story continues with yet newer and faster processors and memory devices.

When one firm, like Intel, is the sole producer of a good, it is by definition a monopoly in the market for that good. An important feature of today's economy is that monopolies, as in the computer chip example, frequently do not last very long. Rapid changes in technology can make patents and copyrights useless well before their life is over.

Nevertheless, some monopolies still last a long time. De Beers is one of the most famous examples of a monopoly. It controls 80 percent of the world's diamond supply and, therefore, is virtually a monopoly. It has maintained its monopoly position since 1929.

Monopolies operate very differently from firms in competitive markets. The biggest difference is that monopolies have the power to set the price in their markets. They use this power to charge higher prices than competitive firms would.

The aim of this chapter is to develop a model of monopoly that can be used to explain this behavior and thereby help us understand how real-world monopolies operate. The model explains how a monopoly decides what price to charge its customers and what quantity to sell. It shows that monopolies cause a loss to society when compared with firms providing goods in competitive markets; the model also provides a way to measure that loss. We also use the model to explain some puzzling pricing behavior, such as why some airlines charge a lower fare to travelers who stay over on a Saturday night.

Monopolies and the reasons for their existence raise important public policy questions about the role of government in the economy. For example, the loss that monopolies cause to society creates a potential role for government: It may step in to reduce this loss.

A Model of Monopoly

monopoly: one firm in an industry selling a product for which there are no close substitutes.

barriers to entry: anything that prevents firms from entering a market.

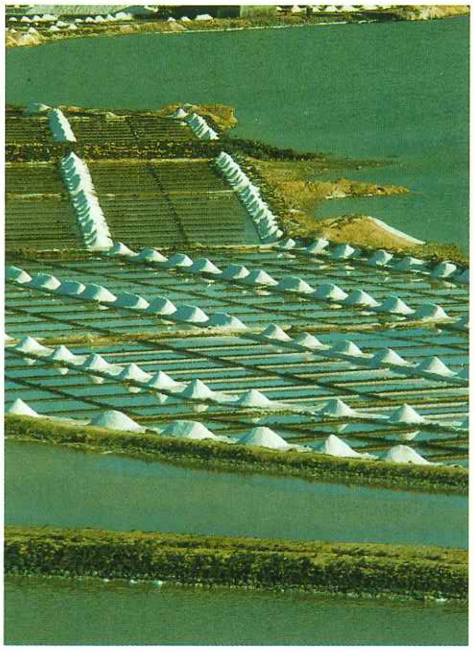
A **monopoly** occurs when there is only one firm in an industry selling a product for which there are no close substitutes. Thus, implicit in the definition of monopoly are **barriers to entry**—other firms are not free to enter the industry. For example, The Diamond Trading Company creates barriers to entry by maintaining exclusive rights to the diamonds in most of the world's diamond mines.

The economist's model of a monopoly assumes that the monopoly will choose a level of output that maximizes profits. In this respect, the model of a monopoly is like that of a competitive firm. If increasing production will increase a monopoly's profits, then the monopoly will raise production, just as a competitive firm would. If cutting production will increase a monopoly's profits, then the monopoly will cut its production, just as a competitive firm would.

market power: a firm's power to set its price without losing its entire share of the market.

price-maker: a firm that has the power to set its price, rather than taking the price set by the market.

The difference between a monopoly and a competitive firm is not what motivates them, but rather how their actions affect the market price. The most important difference is that a monopoly has **market power.** That is, a monopoly has the power to set the price in the market, whereas a competitor does not. This is why a monopoly is called a **price-maker** rather than a *price-taker*, the term used to refer to a competitive firm.



The monopoly power of the salt industry has been illustrated throughout history. Salt monopolies have contributed to the rise of several state powers—from governments in ancient China to medieval Europe, where Venice's control of the salt monopoly helped finance their navy and allowed them to dominate world trade.

Getting an Intuitive Feel for the Market Power of a Monopoly

We can demonstrate the monopoly's power to affect the price in the market by looking at either what happens when the monopoly changes its price or what happens when the monopoly changes the quantity it produces. We consider the price decision first.

There Is No One to Undercut the Monopolist's Price. When there are several sellers competing with one another in a competitive market, one seller can try to sell at a higher price, but no one will buy at that price because there is always another seller nearby who will undercut that price. If a seller charges a higher price, everyone will ignore that seller; there is no effect on the market price.

The monopoly's situation is quite different. Instead of there being several sellers, there is only one seller. If the single seller sets a high price, it has no need to worry about being undercut by other sellers. There are no other sellers. Thus, the single seller—the monopoly—has the power to set a high price. True, the buyers will probably buy less at the higher price—that is, as the price rises, the quantity demanded declines—but because there are no other sellers, they will probably buy something from the lone seller.

The Impact of Quantity Decisions on the Price. Another way to see this important difference between a monopoly and a competitor is to examine what happens to the price when a firm changes the quantity it produces. Suppose that there are 100 firms competing in the bagel-baking market in a large city, each producing about the same quantity of bagels each day. Suppose that one of the firms—Bageloaf—decides to cut its production in half. Although this is a huge cut for one firm, it is a small cut compared to the whole market—only one-half a percent. *Thus the market price will rise very little.* Moreover, if this little price increase affects the behavior of the other 99 firms at all, it will motivate them to increase their production slightly. As they increase the quantity they supply, they partially offset the cut in supply by Bageloaf, and so the change in market price will be even smaller. Thus, by any measure, the overall impact on the price from the change in Bageloaf's production is negligible. Bageloaf has essentially no power to affect the price of bagels in the city.

But now suppose that Bageloaf and the 99 other firms are taken over by Bagelopoly, which then becomes the only bagel bakery in the city. Now, if Bagelopoly cuts production in half, the total quantity of bagels supplied to the whole market is cut in half, and *this will have a big effect on the price in the market.*

If Bagelopoly cut its production even further, the price would rise further. However, if Bagelopoly increased the quantity it produced, the price of bagels would fall. Thus, Bagelopoly has immense power to affect the price. Even if Bagelopoly does not know exactly what the price elasticity of demand for bagels is, it can adjust the quantity it will produce either up or down in order to change the price.

Showing Market Power with a Graph. Figure 10.1 contrasts the market power of a monopoly with that of a competitive firm. The right-hand graph shows that the competitive firm views the market price as essentially out of its control. The market price is shown by the flat line and is thus the same regardless of how much the firm produces. If the competitive firm tried to charge a higher price, nobody would buy because there would be many competitors charging a lower price; so, effectively, the competitive firm cannot charge a higher price.

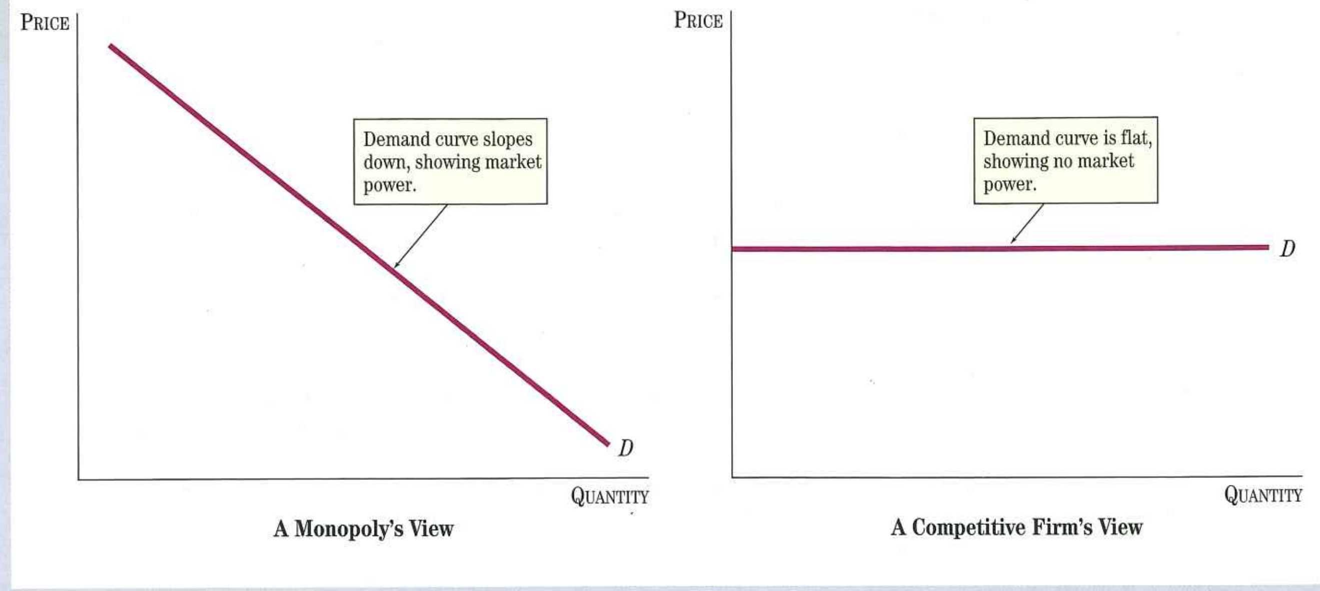


Figure 10.1
How the Market Power of a Monopoly and a Competitive Firm Differ

A monopoly is the only firm in the market. Thus, the market demand curve and the demand curve of the monopoly are the same. By raising the price, the monopoly sells less. In contrast, the competitive firm has no impact on the market price. If the competitive firm charges a higher price, its sales will drop to zero.

To a monopoly, on the other hand, things look quite different. Because the monopoly is the sole producer of the product, it represents the entire market. The monopoly—shown in the left-hand graph—sees a downward-sloping market demand curve for its product. *The downward-sloping demand curve seen by the monopoly is the same as the market demand curve.* If the monopoly charges a higher price, the quantity demanded declines along the demand curve. With a higher price, fewer people buy the item, but with no competitors to undercut that higher price, there is still some demand for the product.

The difference in the market power of a monopoly and a competitive firm—illustrated by the slope of the demand curve each faces—causes the difference in the behavior of the two types of firms.

The Effects of a Monopoly's Decision on Revenues

Now that we have seen how the monopoly can affect the price in its market by changing the quantity it produces, let's see how its revenues are affected by the quantity it produces.

Table 10.1 gives a specific numerical example of a monopoly. Depending on the units for measuring the quantity Q, the monopoly could be producing computer chips or diamonds.

The two columns on the left represent the market demand curve, showing that there is a negative relationship between the price and the quantity sold: As the quantity sold rises from 3 to 4, for example, the price falls from \$130 to \$120 per unit.

The third column of Table 10.1 shows what happens to the monopoly's total revenue, or price times quantity, as the quantity of output increases. Observe that at the beginning, when the monopoly increases the quantity produced, total revenue rises:

Table 10.1
Revenue, Costs, and Profits for a Monopoly (price, revenue, and cost measured in dollars)

Market Demand						
Quantity Produced and Sold (Q)	*Price (P)*	**Total Revenue (*TR*)**	**Marginal Revenue (*MR*)**	**Total Costs (*TC*)**	**Marginal Cost (*MC*)**	**Profits**
0	160	0	—	70	—	−70
1	150	150	150	79	9	71
2	140	280	130	84	5	196
3	130	390	110	94	10	296
4	120	480	90	114	20	366
5	110	550	70	148	34	402
6	100	600	50	196	48	404
7	90	630	30	261	65	369
8	80	640	10	351	90	289
9	70	630	−10	481	130	149
10	60	600	−30	656	175	−56

$TR = P \times Q$

$\frac{\text{Change in } TR}{\text{Change in } Q}$

$\frac{\text{Change in } TC}{\text{Change in } Q}$

$TR - TC$

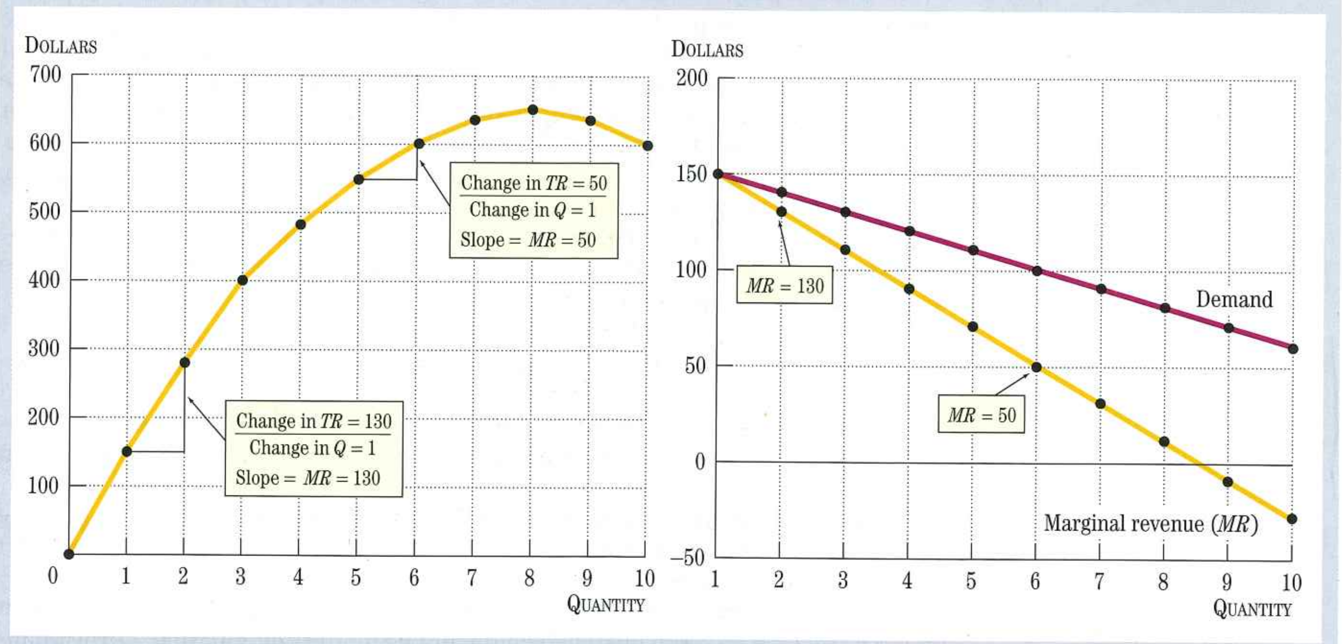


Figure 10.2
Total Revenue, Marginal Revenue, and Demand

The graph on the left plots total revenue for each level of output for Table 10.1. Total revenue first rises and then declines as the quantity of output increases. Marginal revenue is the change in total revenue for each additional increase in the quantity of output and is shown by the yellow curve at the right. Observe that the marginal revenue curve lies below the demand curve at each level of output except $Q = 1$.

When zero units are sold, total revenue is clearly zero; when 1 unit is sold, total revenue is 1 × $150, or $150; when 2 units are sold, total revenue is 2 × $140, or $280; and so on. However, as the quantity sold increases, total revenue rises by smaller and smaller amounts and eventually starts to fall. In Table 10.1, total revenue reaches a peak of $640 at 8 units sold and then starts to decline.

The left-hand graph in Figure 10.2 shows how total revenue changes with the quantity of output for the example in Table 10.1. It shows that total revenue reaches a maximum. Although a monopolist has the power to influence the price, this does not mean that it can get as high a level of total revenue as it wants. Why does total revenue increase by smaller and smaller amounts and then decline as production increases? Because in order to sell more output, the monopolist must lower the price in order to get people to buy the increased output. As it raises output, it must lower the price more and more, and this causes the increase in total revenue to get smaller. As the price falls to very low levels, revenue actually declines.

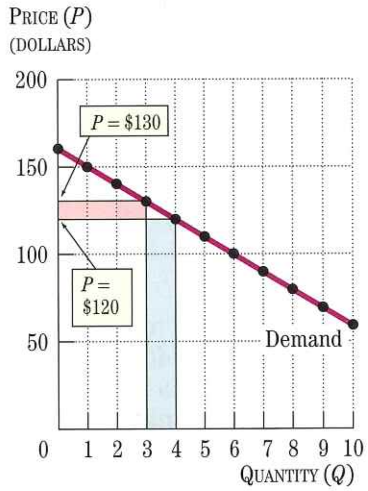
Declining Marginal Revenue. In order to determine the quantity the monopolist produces to maximize profits, we must measure marginal revenue. *Marginal revenue*, introduced in Chapter 6, is the change in total revenue from one more unit of output sold. For example, if total revenue increases from $480 to $550 as output rises by 1 unit, marginal revenue is $70 ($550 − $480 = $70). Marginal revenue for the monopolist in Table 10.1 is shown in the fourth column, next to total revenue. In addition, marginal revenue is plotted in the right-hand graph of Figure 10.2, where it is labeled *MR*.

An important relationship between marginal revenue and price, shown in Table 10.1 and Figure 10.2, is that *marginal revenue declines as the quantity of output rises.* This is just another way to say that the changes in total revenue get smaller and smaller as output increases, as we already noted.

Marginal Revenue Is Less than the Price. Another important relationship between marginal revenue and price is that for every level of output, *marginal revenue is less than the price* (except at the first unit of output, where it equals the price). To observe this, compare the price (*P*) and marginal revenue (*MR*) in Table 10.1 or in the right-hand panel of Figure 10.2.

Note that the red line in Figure 10.2 showing the price and the quantity of output demanded is simply the demand curve facing the monopolist. Thus, another way to say that marginal revenue is less than the price at a given level of output is to say that the *marginal revenue curve lies below the demand curve.*

Why is the marginal revenue curve below the demand curve? When the monopolist increases output by one unit, there are two effects on total revenue: (1) a positive effect, which equals the price *P* times the additional unit sold, and (2) a negative effect, which equals the reduction in the price on all items previously sold times the number of such items sold. For example, as the monopolist in Table 10.1 increases production from 4 to 5 units and the price falls from $120 to $110, marginal revenue is $70; this $70 is equal to the increased revenue from the extra unit produced, or $110, less the decreased revenue from the reduction in the price, or $40 ($10 times the 4 units previously produced). Marginal revenue (*MR* = $70) is thus less than the price (*P* = $110). The two effects on marginal revenue are shown in the graph in the margin when quantity increases from 3 to 4. Other numerical examples are shown in the table on page 262. Because the second effect—the reduction in revenue due to the lower price on the items previously produced—is subtracted from the first, the price is always greater than the marginal revenue.



Graph Showing the Two Effects on Marginal Revenue
When the monopoly raises output from 3 units to 4 units, revenue increases; that is, marginal revenue is greater than zero. There is a positive effect (blue rectangle) because one more item is sold and a negative effect (red rectangle) because prices on previously sold items fall. Here the positive effect (area of blue rectangle is $120) is greater than the negative effect (area of red rectangle is $30), so marginal revenue is $90.

Marginal Revenue Can Be Negative. Note that marginal revenue is negative when output is 9 or 10 units in the example. Then total revenue *falls* as additional units are produced. It would be crazy for a monopolist to produce so much that its marginal revenue was negative.

Marginal revenue is negative when the price elasticity of demand is less than 1. To see this, some algebra is helpful. Note from the examples in the table below that the following equation holds.

$$MR = (P \times \Delta Q) - (\Delta P \times Q)$$

If $MR < 0$, then

$$P \times \Delta Q < \Delta P \times Q$$

which implies that

$$\frac{(\Delta Q/Q)}{(\Delta P/P)} < 1$$

or, in words, that the price elasticity of demand is less than 1. Thus, we conclude that a monopoly would never produce a level of output where the price elasticity of demand is less than 1.

Quantity Sold	Marginal Revenue (*MR*)		Price × (Change in Quantity)	−	(Change in Price) × (Previous Quantity Sold)
			P × (Δ*Q*)		(Δ*P*) × (*Q*)
1	150	=	$150 × 1	−	$10 × 0
2	130	=	$140 × 1	−	$10 × 1
3	110	=	$130 × 1	−	$10 × 2
4	90	=	$120 × 1	−	$10 × 3
5	70	=	$110 × 1	−	$10 × 4
6	50	=	$100 × 1	−	$10 × 5
7	30	=	$90 × 1	−	$10 × 6
8	10	=	$80 × 1	−	$10 × 7
9	−10	=	$70 × 1	−	$10 × 8
10	−30	=	$60 × 1	−	$10 × 9

average revenue: total revenue divided by quantity.

Average Revenue. We can also use average revenue to show that marginal revenue is less than the price. **Average revenue** is defined as total revenue divided by the quantity of output; that is, $AR = TR/Q$. Because total revenue (TR) equals price times quantity ($P \times Q$), we can write average revenue (AR) as $(P \times Q)/Q$ or, simply, the price P. In other words, the demand curve—which shows price at each level of output—also shows average revenue for each level of output.

Now recall from Chapter 8 that when the average of anything (costs, grades, heights, or revenues) declines, the marginal must be less than the average. Thus, because average revenues decline (that is, the demand curve slopes down), the marginal revenue curve must lie below the demand curve.

Finding Output to Maximize Profits at the Monopoly

Now that we have seen how a monopoly's revenues depend on the quantity it produces, let's see how its profits depend on the quantity it produces. Once we identify the relationship between profit and the quantity the monopoly will produce, we can determine the level of output that maximizes the monopoly's profits. To determine profits, we must look at the costs of the monopoly and then subtract total costs from total revenue.

Observe that the last three columns of Table 10.1 on page 260 show the costs and profits for the example monopoly. There are no new concepts about a monopoly's costs compared to a competitive firm's costs, so we can use the cost measures we developed in Chapters 7 to 9. The most important concepts are that total costs increase as more is produced and that marginal cost also increases, at least for high levels of output.

Comparing Total Revenue and Total Costs. The difference between total revenue and total costs is profits. Observe in Table 10.1 that as the quantity produced increases, both the total revenue from selling the product and the total costs of producing the product increase. However, at some level of production, total costs start to increase more than revenue increases, so that eventually profits must reach a maximum.

A quick glance at the profits column in Table 10.1 will show that this maximum level of profits is $404 and is reached when the monopoly produces 6 units of output. The price the monopoly must charge so that people will buy 6 units of output is $100, according to the second column of Table 10.1.

To help you visualize how profits change with quantity produced and to find the maximum level of profits, Figure 10.3 plots total costs, total revenue, and profits from Table 10.1. Profits are shown as the gap between total costs and total revenue. The gap reaches a maximum when output Q equals 6.

Equating Marginal Cost and Marginal Revenue. There is an alternative, more intuitive approach to finding the level of production that maximizes a monopolist's profits. This approach looks at marginal revenue and marginal cost and employs a rule that economists use extensively.

Consider producing different levels of output, starting with 1 unit and then rising unit by unit. Compare the marginal revenue from selling each additional unit of output with the marginal cost of producing it. If the marginal revenue is greater than

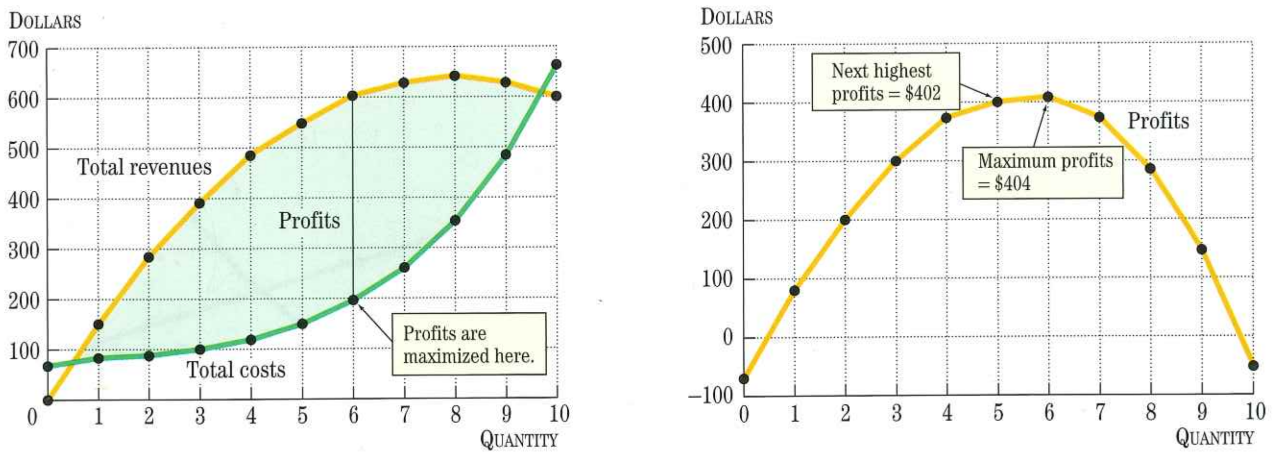


Figure 10.3
Finding a Quantity of Output to Maximize Profits

Profits are shown as the gap between total revenue and total costs in the graph on the left and are plotted on the graph on the right. Profits are at a maximum when the quantity of output is 6.

the marginal cost of the additional unit, then profits will increase if the unit is produced. Thus, the unit should be produced, because total revenue rises by more than total costs. For example, in Table 10.1, the marginal revenue from producing 1 unit of output is $150 and the marginal cost is $9. Thus, at least 1 unit should be produced. What about 2 units? Then marginal revenue equals $130 and marginal cost equals $5, so it makes sense to produce 2 units. Continuing this way, the monopolist should increase its output as long as marginal revenue is greater than marginal cost. But because marginal revenue is decreasing and eventually marginal cost is increasing, at some level of output marginal revenue will drop below marginal cost. The monopolist should not produce at that level. For example, in Table 10.1, the marginal revenue from selling 7 units of output is less than the marginal cost of producing it. Thus, the monopolist should not produce 7 units; instead, 6 units of production, with $MR = 50$ and $MC = 48$, is the profit-maximization level; this is the highest level of output for which marginal revenue is greater than marginal cost. Note that this level of output is exactly what we obtain by looking at the gap between total revenue and total costs.

Thus, *the monopolist should produce up to the level of production where marginal cost equals marginal revenue* ($MC = MR$). If the level of production cannot be adjusted so exactly that marginal revenue is precisely equal to marginal cost, then the firm should produce at the highest level of output for which marginal revenue exceeds marginal cost, as in Table 10.1. In most cases, the monopoly will be able to adjust its output by smaller fractional amounts (for example, pounds of diamonds rather than tons of diamonds), and therefore marginal revenue will equal marginal cost.

A picture of how this marginal revenue equals marginal cost rule works is shown in Figure 10.4. The marginal revenue curve is plotted, along with the marginal cost curve. As the quantity produced increases above very low levels, the marginal cost

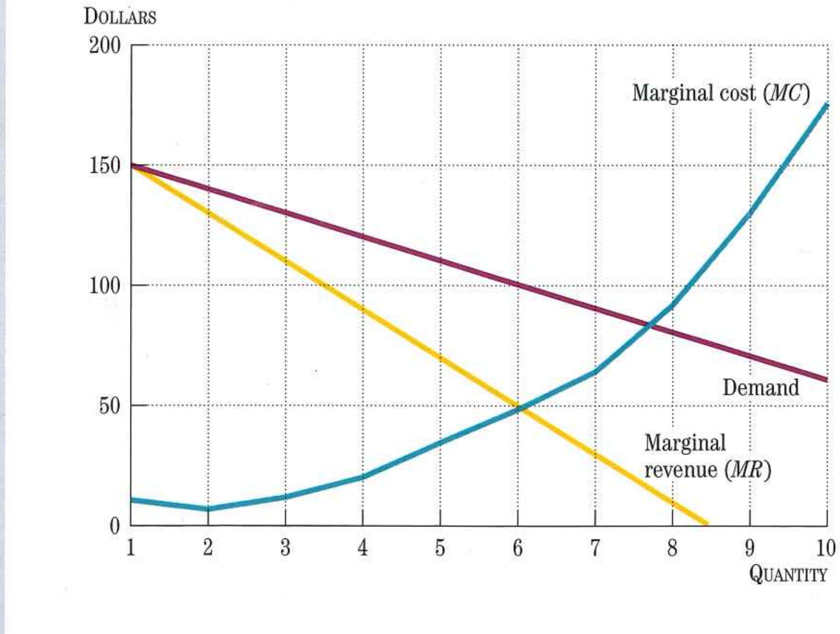


Figure 10.4
Marginal Revenue and Marginal Cost
The profit-maximizing monopoly will produce up to the point where marginal revenue equals marginal cost, as shown in the diagram. If fractional units cannot be produced, then the monopoly will produce at the highest level of output for which marginal revenue is greater than marginal cost. These curves are drawn for the monopoly in Table 10.1.

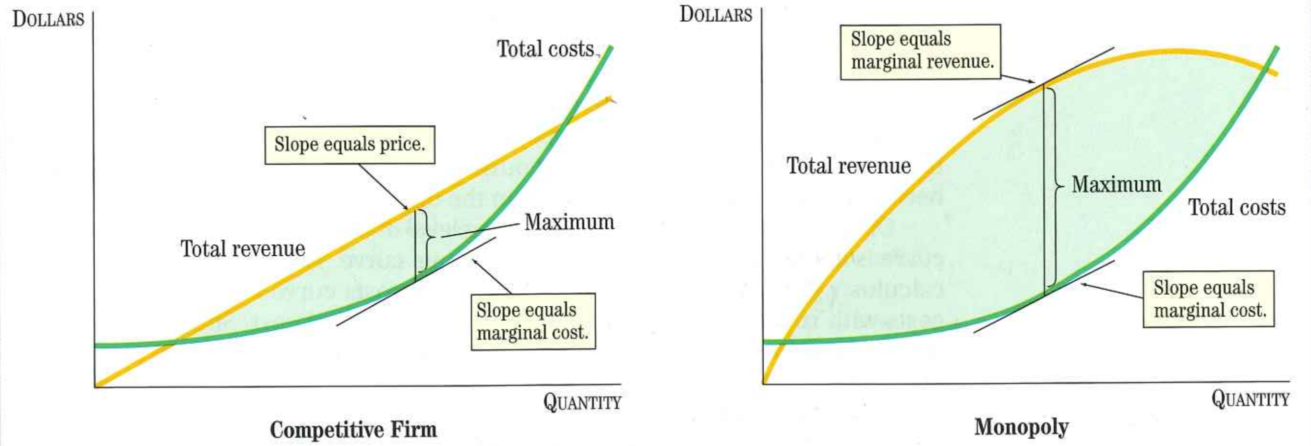


Figure 10.5
Profit Maximization for a Monopoly and a Competitive Firm

Total revenue for a competitive firm rises steadily with the amount sold; total revenue for a monopoly first rises and then falls. However, both the monopolist and the competitor maximize profits by making the gap between the total costs curve and the total revenue curve as large as possible or by setting the slope of the total revenue curve equal to the slope of the total costs curve. Thus, marginal revenue equals marginal cost. For the competitive firm, marginal revenue equals the price.

curve slopes up and the marginal revenue curve slopes down. Marginal revenue equals marginal cost at the level of output where the two curves intersect.

MC = *MR* at a Monopoly versus *MC* = *P* at a Competitive Firm

It is useful to compare the *MC* = *MR* rule for the monopolist with the *MC* = *P* rule for the competitive firm that we derived in Chapter 6.

MR=P
MC=MR

Marginal Revenue Equals the Price for a Price-Taker. For a competitive firm, total revenue is equal to the quantity sold (*Q*) multiplied by the market price (*P*), but the competitive firm cannot affect the price. Thus, when the quantity sold is increased by one unit, revenue is increased by the price. In other words, for a competitive firm, marginal revenue equals the price; to say that a competitive firm sets its marginal cost equal to marginal revenue is to say that it sets its marginal cost equal to the price. Thus, the *MC* = *MR* rule applies to both monopolies and competitive firms that maximize profits.

A Graphical Comparison. Figure 10.5 is a visual comparison of the two rules. A monopoly is shown on the right graph of Figure 10.5. This is the kind of graph we drew in Figure 10.3 except that it applies to any firm, so we do not show the units. A competitive firm is shown in the left graph of Figure 10.5. This is exactly like the graph showing a competitive firm in Figure 6.8. The scale on these two figures might be quite different; only the shapes are important for this comparison.

Look carefully at the shape of the total revenue curve for the monopoly and contrast it with the total revenue curve for the competitive firm. The total revenue curve for the monopoly starts to turn down at higher levels of output, whereas the total revenue curve for the competitive firm keeps rising in a straight line.

To illustrate the maximization of profits, we have put the same total costs curve on both graphs in Figure 10.5. Both types of firms maximize profits by setting production so that the gap between the total revenue curve and the total costs curve is as large as possible. That level of output, the profit-maximizing level, is shown for both firms. Higher or lower levels of output will reduce profits, as shown by the gaps between total revenue and total costs in the diagrams.

Observe that at the profit-maximizing level of output, the slope of the total costs curve is equal to the slope of the total revenue curve. Those of you who know some calculus will notice that the slope of the total costs curve is the derivative of total costs with respect to production—that is, the marginal cost. Similarly, the slope of the total revenue curve is the marginal revenue—the increase in total revenue when output increases. Thus, we have another way of seeing that marginal revenue equals marginal cost for profit maximization.

For the competitive firm, marginal revenue is the price, which implies the condition of profit maximization at a competitive firm derived in Chapter 6: Marginal cost equals price. However, for the monopolist, marginal revenue and price are not the same thing.

REVIEW

- When one firm is the sole producer of a product with no close substitutes, it is a monopoly. Most monopolies do not last forever. They come and go as technology changes. Barriers to the entry of new firms are needed to maintain a monopoly.
- A monopoly is like a competitive firm in that it tries to maximize profits. But unlike a competitive firm, a monopoly has market power; it can affect the market price. The demand curve the monopoly faces is the same as the market demand curve.
- Marginal revenue is the change in total revenue as output increases by one unit. Marginal revenue is less than the price at each level of output (except the first). If a firm maximizes profits, then marginal revenue equals marginal cost ($MR = MC$).
- For a competitive firm, marginal revenue equals marginal cost equals price ($MR = MC = P$).
- For a monopoly, marginal revenue also equals marginal cost, but marginal revenue does not equal the price. Hence, for the monopolist, price is not necessarily equal to marginal cost.

The Generic Diagram of a Monopoly and Its Profits

Look at Figure 10.6, which combines the monopoly's demand and marginal revenue curves with its average total cost curve and marginal cost curve. This diagram is the workhorse of the model of a monopoly, just as Figure 8.6 on page 210 is the workhorse of the model of a competitive firm. As with the diagram for a competitive firm,

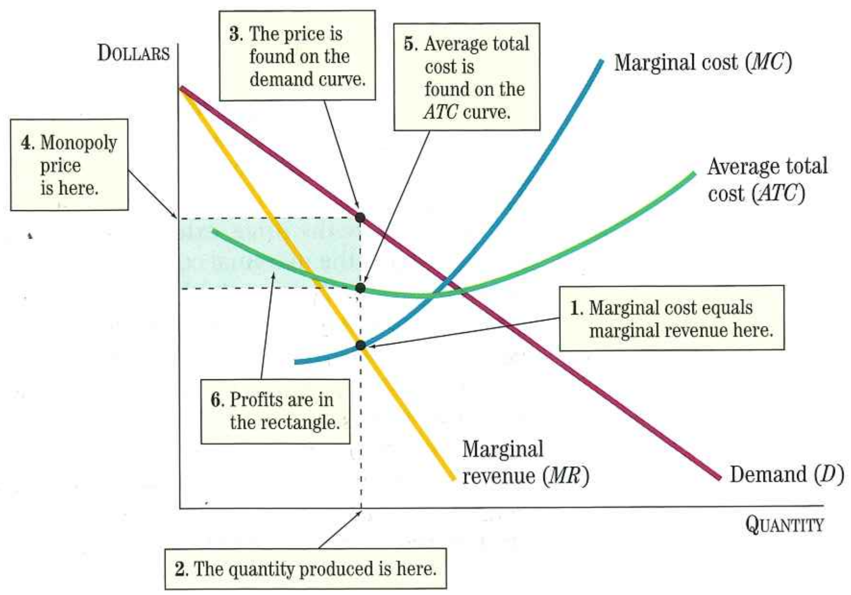


Figure 10.6
The Generic Diagram for a Monopoly
The marginal revenue and demand curves are superimposed on the monopoly's cost curves. The monopoly's production, price, and profits can be seen on the same diagram. Quantity is given by the intersection of the marginal revenue curve and the marginal cost curve. Price is given by the demand curve at the point corresponding to the quantity produced, and average total cost is given by the *ATC* curve at that quantity. Monopoly profits are given by the rectangle that is the difference between total revenue and total costs.

you should be able to draw it in your sleep. It is a generic diagram that applies to any monopolist, not just the one in Table 10.1, so we do not put scales on the axes.

Observe that Figure 10.6 shows four curves: a downward-sloping demand curve (*D*), a marginal revenue curve (*MR*), an average total cost curve (*ATC*), and a marginal cost curve (*MC*). The position of these curves is very important. First, the marginal cost curve cuts through the average total cost curve at the lowest point on the average total cost curve. Second, the marginal revenue curve is below the demand curve over the entire range of production (except at the vertical axis near 1, where they are equal).

We have already given the reasons for these two relationships (in Chapter 8 and in the previous section of this chapter), but it would be a good idea for you to practice sketching your own diagram like Figure 10.6 to make sure the positions of your curves meet these requirements.[1]

Determining Monopoly Output and Price on the Diagram

In Figure 10.6 we show how to calculate the monopoly output and price. First, find the point of intersection of the marginal revenue curve and the marginal cost curve. Second, draw a dashed vertical line through this point and look down the dashed line at the horizontal axis to see what the quantity produced is. Producing a larger quantity would lower marginal revenue below marginal cost. Producing a smaller

[1] When sketching diagrams, it is useful to know that when the demand curve is a straight line, the marginal revenue curve is always twice as steep as the demand curve and, if extended, would cut the horizontal axis exactly halfway between zero and the point where the demand curve would cut the horizontal axis.

quantity would raise marginal revenue above marginal cost. The quantity shown is the profit-maximizing level. It is the amount the monopolist produces.

What price will the monopolist charge? We again use Figure 10.6, but be careful: Unlike the quantity, the monopolist's price is *not* determined by the intersection of the marginal revenue curve and the marginal cost curve. The price has to be such that the quantity demanded is equal to the quantity that the monopolist decides to produce. To find the price, we need to look at the demand curve in Figure 10.6. The demand curve gives the relationship between price and quantity demanded. It tells how much the monopolist will charge for its product in order to sell the amount produced.

To calculate the price, extend the dashed vertical line upward from the point of intersection of the marginal cost curve and the marginal revenue curve until it intersects the demand curve. At the intersection of the demand curve and the vertical line, we find the price that will generate a quantity demanded equal to the quantity produced. Now draw a horizontal line over to the left from the point of intersection to mark the price on the vertical axis. This is the monopoly's price, about which we will have more to say later.

Determining the Monopoly's Profits

Profits can also be shown on the diagram in Figure 10.6. Profits are given by the difference between the area of two rectangles, a total revenue rectangle and a total costs rectangle. Total revenue is price times quantity and is thus equal to the area of the rectangle with height equal to the monopoly price and length equal to the quantity produced. Total costs are average total cost times quantity and are thus equal to the area of the rectangle with height equal to *ATC* and length equal to the quantity produced. Profits are then equal to the green-shaded area that is the difference between these two rectangles.

It is possible for a monopoly to have negative profits, or losses, as shown in Figure 10.7. In this case, the price is below average total cost, and therefore total

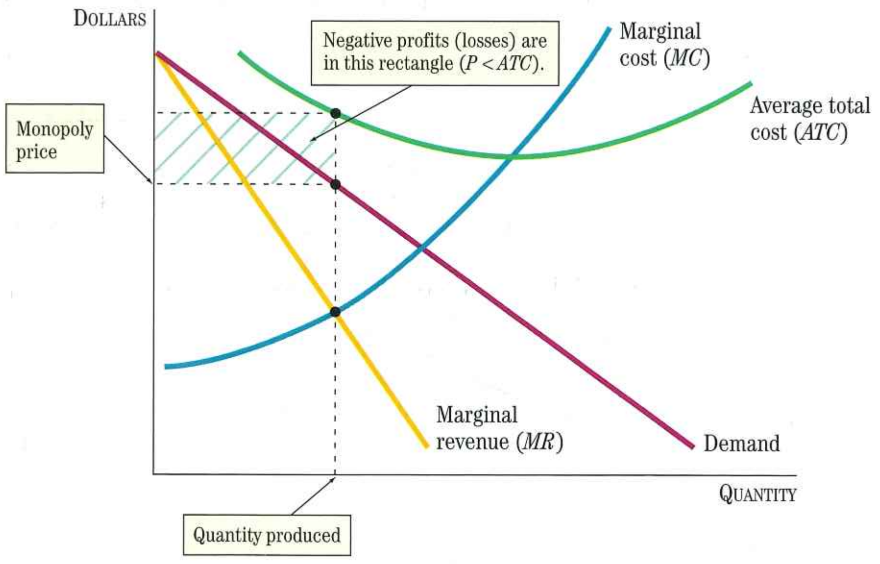


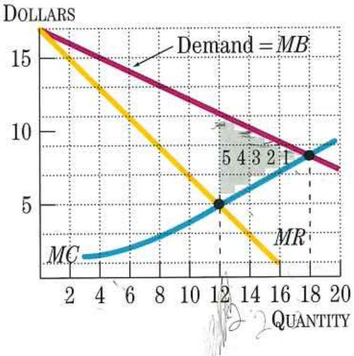
Figure 10.7
A Monopoly with Negative Profits
If a monopoly finds that average total cost is greater than the price at which marginal revenue equals marginal cost, then it runs losses. If price is also less than average variable cost, then the monopoly should shut down, just like a competitive firm.

revenue is less than total costs. Like a competitive firm, a monopolist with negative profits will shut down if the price is less than average variable cost. It will eventually exit the market if negative profits persist.

REVIEW

- A monopolist's profit-maximizing output and price can be determined graphically. The diagram shows four curves: the marginal revenue curve, the demand curve, the marginal cost curve, and the average total cost curve. The marginal revenue curve is always below the demand curve.
- The monopoly's production is determined at the point where marginal revenue equals marginal cost.
- The monopoly's price is determined from the demand curve at the point where the quantity produced equals the quantity demanded.
- The monopoly's profits are determined by subtracting the total costs rectangle from the total revenue rectangle. The total revenue rectangle is given by the price times the quantity produced. The total costs rectangle is given by the average total cost times the quantity produced.

Competition, Monopoly, and Deadweight Loss



Numerical Example of Deadweight Loss Calculation
The monopoly shown in the diagram above produces only 12 items, but a competitive industry would produce 18 items. For the 13th through 17th items, which are not produced by the monopoly, the marginal benefit is greater than the marginal cost by the amounts $5, $4, $3, $2, and $1, respectively, as shown by the areas between the demand curve and the supply curve for the competitive industry. Hence, the deadweight loss caused by the monopoly is the sum $5 + $4 + $3 + $2 + $1 = $15.

Are monopolies harmful to society? Do they reduce consumer surplus? Can we measure these effects? To answer these questions, economists compare the price and output of a monopoly with those of a competitive industry. First, observe in Figure 10.6 or Figure 10.7 that the monopoly does not operate at the minimum point on the average total cost curve even in the long run. Recall that firms in a competitive industry do operate at the lowest point on the average total cost curve in the long run.

To go further in our comparison, we use Figure 10.8, which is a repeat of Figure 10.6, except that the average total cost curve is removed to reduce the clutter. All the other curves are the same.

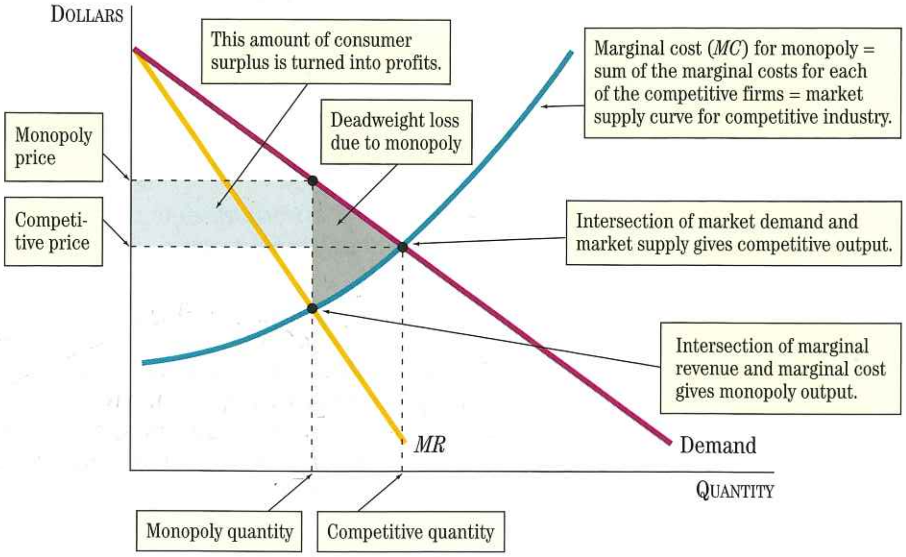
Comparison with Competition

Suppose that instead of there being only one firm in the market, there are now many competitive firms. For example, suppose Bagelopoly—a single firm producing bagels in a large city—is broken down into 100 different bagel bakeries like Bageloaf. The production point for the monopolistic firm and its price before the breakup are marked as "monopoly quantity" and "monopoly price" in Figure 10.8. What are production and price after the breakup?

The market supply curve for the new competitive industry would be Bagelopoly's old marginal cost curve because this is the sum of the marginal cost curves of all the newly created firms in the industry. Equilibrium in the competitive industry is where this market supply curve crosses the market demand curve. The amount of production at that point is marked by "competitive quantity" in Figure 10.8. The price at that equilibrium is marked by "competitive price" on the vertical axis.

Compare the quantity and price for the monopolist and the competitive industry. It is clear from the diagram in the margin that the quantity produced by the monopolist is less than the quantity produced by the competitive industry. It is also

Figure 10.8
Deadweight Loss from Monopoly
The monopolist's output and price are determined as in Figure 10.6. To get the competitive price, we imagine that competitive firms make up an industry supply curve that is the same as the monopolist's marginal cost curve. The competitive price and quantity are given by the intersection of the supply curve and the demand curve. The monopoly quantity is lower than the competitive quantity. The monopoly price is higher than the competitive price. The deadweight loss is the reduction in consumer plus producer surplus due to the lower level of production by the monopolist.



clear that the monopoly will charge a higher price than will emerge from a competitive industry. In sum, the monopoly produces less and charges a higher price than the competitive industry would.

This is a very important result. The monopoly exploits its market power by holding back on quantity produced and causing the price to rise compared with the competitive equilibrium. This is always the case. Convince yourself by drawing different diagrams. For example, when De Beers exercises its market power, it holds back production of diamonds, thereby raising the price and earning economic profits.

Note that even though the monopoly has the power to do so, it does not increase its price without limit. When the price is set very high, marginal cost rises above marginal revenue. That behavior is not profit-maximizing.

Deadweight Loss from Monopoly

The economic harm caused by a monopoly occurs because it produces less than a competitive industry would. How harmful, then, is a monopoly?

Consumer Surplus and Producer Surplus Again. Economists measure the harm caused by monopolies by the decline in the sum of consumer surplus plus producer surplus. Recall that *consumer surplus* is the area above the market price line and below the demand curve, the demand curve being a measure of consumers' marginal benefit from consuming the good. The *producer surplus* is the area above the marginal cost curve and below the market price line. Consumer surplus plus producer surplus is thus the area between the demand curve and the marginal cost curve. It measures the sum of the marginal benefits to consumers of the good less the sum of the marginal costs to the producers of the good. A competitive market will maximize the sum of consumer plus producer surplus.

With a lower quantity produced by a monopoly, however, the sum of consumer surplus and producer surplus is reduced, as shown in Figure 10.8. This reduction in consumer plus producer surplus is called the *deadweight loss due to monopoly*. It is a quantitative measure of the harm a monopoly causes the economy. A numerical example is given in the margin.

How large is the deadweight loss in the U.S. economy? Using the method illustrated in Figure 10.8, empirical economists estimate that the loss is between .5 and 2 percent of GDP, or between $60 billion and $240 billion per year. Of course, the deadweight loss is a larger percentage of production in industries where monopolies are a greater presence.

Figure 10.8 also shows that the monopoly takes, in the form of profits, some of the consumer surplus that would have gone to the consumers in competitive markets. Consumer surplus is now the area below the demand curve and above the monopoly price, which is higher than the competitive price. However, this transfer of consumer surplus to the monopoly is not a deadweight loss, because what the consumers lose, the monopoly gains. This transfer affects the distribution of income, but it is not a net loss to society.

■ **Meaningful Comparisons.** In any given application, one needs to be careful that the comparison of monopoly and competition makes sense. Some industries cannot be broken up into many competitive firms the way bagel bakeries can. Having 100 water companies serving one local area, for example, would be very costly. The choice for society is not between a monopolistic water industry and a competitive water industry. Although we might try to affect the monopoly's decisions by government actions, we should not try to break up water services within each community.

History provides many other examples. Western settlers in the United States during the nineteenth century had a larger consumer surplus from the monopolist railroads—in spite of the monopolists' profits—than they did from competitive wagon trains. Modern-day users of the information highway—computers and telecommunications—reap a larger consumer surplus from Intel's computer chips, even if they are produced monopolistically, than they would from a competitive abacus industry.

The Monopoly Price Is Greater than Marginal Cost

Another way to think about the loss to society from monopoly is to observe the difference between price and marginal cost. Figure 10.8, for example, shows that the monopoly price is well above the marginal cost at the quantity where the monopoly chooses to produce.

■ **Marginal Benefit Is More than Marginal Cost.** Because consumers will consume up to the point where the marginal benefit of a good equals its price, the excessive price means that the marginal benefit of a good is greater than the marginal cost. This is inefficient because producing more of the good would increase benefits to consumers by more than the cost of producing it.

The size of the difference between price and marginal cost depends on the elasticity of the monopoly's demand curve. If the demand curve is highly elastic (close to a competitive firm's view as shown in Figure 10.1), then the difference between price and marginal cost is small.

price-cost margin: the difference between price and marginal cost divided by the price. This index is an indicator of market power, where an index of 0 indicates no market power and a higher price-cost margin indicates greater market power.

The Price-Cost Margin. A common measure of the difference between price and marginal cost is the **price-cost margin.** It is defined as

$$\frac{\text{Price minus marginal cost}}{\text{Price}}$$

For example, if the price is \$4 and the marginal cost is \$2, the price-cost margin is $(4 - 2)/4 = .5$. The price-cost margin for a competitive firm is zero because price equals marginal cost.

Economists use a rule of thumb to show how the price-cost margin depends on the price elasticity of demand. The rule of thumb is shown in the equation below.

$$\text{Price-cost margin} = \frac{1}{\text{price elasticity of demand}}$$

For example, when the elasticity of demand is 2, the price-cost margin is .5. The flat demand curve has an infinite elasticity, in which case the price-cost margin is zero; in other words, price equals marginal cost.

REVIEW

- A monopoly creates a deadweight loss because it restricts output below what the competitive market would produce. The cost is measured by the deadweight loss, which is the reduction in the sum of consumer plus producer surplus.
- Sometimes the comparison between monopoly and competition is only hypothetical because it would either be impossible or make no sense to break up the monopoly into competitive firms.
- Another way to measure the impact of a monopoly is by the difference between price and marginal cost. Monopolies always charge a price higher than marginal cost. The difference—summarized in the price-cost margin—depends inversely on the elasticity of demand.

Why Monopolies Exist

Given this demonstration that monopolies lead to high prices and a deadweight loss to society, you may be wondering why monopolies exist. In this section, we consider three reasons for the existence of monopolies.

Natural Monopolies

The nature of production is a key factor in determining the number of firms in the industry. If big firms are needed in order to produce at low cost, it may be natural for a few firms or only one firm to exist. In particular, *economies of scale*—a declining long-run average total cost curve over some range of production—can lead to a monopoly. Recall from Chapter 8 that the *minimum efficient scale* of a firm is the minimum size of the firm for which average total costs are lowest. If the minimum efficient scale is only a small fraction of the size of the market, then there will be many firms.

For example, suppose the minimum efficient scale for beauty salons in a city is a size that serves 30 customers a day at each salon. Suppose the quantity of hair stylings demanded in the city is 300 per day. We can then expect there will be 10 beauty salons (300/30 = 10) in the city. But if the minimum efficient scale is larger (for example, 60 customers per day), then the number of firms in the industry will be smaller (for example, 300/60 = 5 salons). At the extreme case where the minimum efficient scale of the firm is as large as or larger than the size of the market (for example, 300 per day), there will probably be only one firm (300/300 = 1), which will be a monopoly.

A water company in a small town, for example, has a minimum efficient scale larger than the number of houses and businesses in the town. There are huge fixed costs to lay pipe down the street, but each house connection has a relatively low cost, so average total cost declines as more houses are connected. Other industries that usually have a very large minimum efficient scale are electricity and local telephone service. In each of these industries, average total cost is lowest if one firm delivers the service. **Natural monopolies** exist when average total cost is declining and the minimum efficient scale is larger than the size of the market.

natural monopoly: a single firm in an industry in which average total cost is declining over the entire range of production and the minimum efficient scale is larger than the size of the market.

The prices charged by many natural monopolies are regulated by government. The purpose of the regulation is to keep the price below the monopoly price and closer to the competitive price. Such regulation can thereby reduce the deadweight loss of the monopoly. Alternative methods of regulating natural monopolies are discussed in Chapter 12. Water companies and electric companies are regulated by government.

A change in technology that changes the minimum efficient scale of firms can radically alter the number of firms in the industry. For example, AT&T used to be viewed as a natural monopoly in long-distance telephone service. Because laying copper wire across the United States required a huge cost, it made little sense to have more than one firm. The U.S. government regulated the prices that AT&T charged its customers, endeavoring to keep the price of calls below the monopoly price and closer to the competitive price. But when the technology for transmitting signals by microwave developed, it became easier for other firms also to provide services. Thus, MCI and Sprint, as well as AT&T, could provide services at least as cheaply as one firm. Because of this technological change, the government decided to end the AT&T monopoly by allowing MCI and Sprint to compete with AT&T. Nationwide telephone service is no longer a monopoly.

Patents and Copyrights

Another way monopolies arise is through the granting of patents and copyrights by the government. Intel's patent was the source of its monopoly on its computer chips. The U.S. Constitution and the laws of many other countries require that government grant patents to inventors. If a firm registers an invention with the U.S. government, it can be granted a monopoly in the production of that item for 20 years. In other words, the government prohibits other firms from producing the good without the permission of the patent holder. Patents are given for many inventions, including the discovery of new drugs. Pharmaceutical companies hold patents on many of their products, giving them a monopoly to produce and sell these products. Copyrights on computer software, chips, movies, and books also give firms the sole right to market the products. Thus, patents and copyrights can create monopolies.

The award of monopoly rights through patents and copyrights serves a useful purpose. It can stimulate innovation by rewarding the inventor. In other words, the chance to get a patent or copyright gives the inventor more incentive to devote time and resources to invent new products or to take a risk and try out new ideas.

Pharmaceutical companies, for example, argue that their patents on drugs are a reward for inventing the drugs. The higher prices and deadweight loss caused by the patent can be viewed as the cost of the new ideas and products. By passing laws to control drug prices, government could lower the prices of drugs to today's sick people. This would be popular, but doing so would reduce the incentive for the firms to invent new drugs. Society—and, in particular, people in future years—could suffer a loss. When patents expire, we usually see a major shift toward competition. In general, when assessing the deadweight loss due to monopoly, one must consider the benefits of the research and the new products that monopoly profits may create.

As technology has advanced, patents and copyrights have had to become increasingly complex in order to prevent firms from getting around them. Nevertheless, patent and copyright protection does not always work in maintaining the monopoly. Many times potential competing firms get around copyrights on computer software and chips by "reverse engineering," in which specialists look carefully at how each part of a product works, starting with the final output. Elaborate mechanisms have been developed, such as "clean rooms," in which one group of scientists and programmers tells another group what each subfunction of the invention does but does not tell them how it is done. The other group then tries to invent an alternative way to perform the task. Because they cannot see how it is done, they avoid violating the copyright.

Licenses

Sometimes the government creates a monopoly by giving or selling to one firm a license to produce the product. The U.S. Postal Service is a government-sponsored monopoly. A law makes it illegal to use a firm other than the U.S. Postal Service for first-class mail. However, even this monopoly is diminishing with competition from overnight mail services and fax technology.

National parks sometimes grant or sell to single firms licenses to provide food and lodging services. The Curry Company, for example, was granted a monopoly to provide services in Yosemite National Park. For a long time the Pennsylvania Turnpike—a toll road running the width of the state—licensed a monopoly to Howard Johnson Company to provide food for travelers on the long stretches of the road. In recent years, seeing the advantage of competition, the turnpike authorities have allowed several different fast-food chains to get licenses.

Attempts to Monopolize and Erect Barriers to Entry

Adam Smith warned that firms would try to create monopolies in order to raise their prices. One of the reasons Smith favored free trade between countries was that it would reduce the ability of firms in one country to form a monopoly; if they did form a monopoly and there were no restrictions on trade serving as barriers to entry for firms in other countries, then foreign firms would break the monopoly.

History shows us many examples of firms attempting to monopolize an industry by merging with other firms and then erecting barriers to entry. De Beers is one example of such a strategy apparently being successful on a global level. In the last part of the nineteenth century, several large firms were viewed as monopolies. Standard Oil, started by John D. Rockefeller in the 1880s, is a well-known example. The firm had control of most of the oil-refining capacity in the United States. Thus, Standard Oil was close to having a monopoly in oil refining. However, the federal government forced Standard Oil to break up into smaller firms. We will consider other examples of the government's breaking up monopolies or preventing them from forming in Chapter 12.

Barriers to entry allow a monopoly to persist, so for a firm to maintain a monopoly, it needs barriers to entry. The box "Reading the News About Barriers to Entry in the Microprocessor Industry" discusses a recent suit that has been brought by Advanced Micro Devices against Intel for erecting barriers to entry.

Barriers to entry can also be created by professional certification. For example, economists have argued that the medical and legal professions in the United States erect barriers to the entry of new doctors and lawyers by having tough standards for admittance to medical school or to the bar and by restricting the types of services that can be performed by nurses or paralegals. Doctors' and lawyers' fees might be lower if there were lower barriers to entry and, therefore, more competition.

Simply observing that a firm has no competitors is not enough to prove that there are barriers to entry. Sometimes the threat of potential entry into a market may be enough to get a monopolist to act like a competitive firm. For example, the possibility of a new bookstore's opening up off campus may put pressure on the campus bookstore to keep its prices low. When other firms, such as off-campus bookstores, can potentially and easily enter the market, they create what economists call a **contestable market.** In general, the threat of competition in contestable markets can induce monopolists to act like competitors.

contestable market: a market in which the threat of competition is enough to encourage firms to act like competitors.

REVIEW

- Economies of scale, patents, copyrights, and licenses are some of the reasons monopolies exist.
- Natural monopolies are frequently regulated by government.
- Many large monopolies in the United States, such as Standard Oil and AT&T, have been broken apart by government action.

Price Discrimination

In the model of monopoly we have studied in this chapter, the monopolists charge a single price for the good they sell. In some cases, however, firms charge different people different prices for the same item. This is called **price discrimination.** Price discrimination is common and is likely to become more common in the future as firms become more sophisticated in their price setting. Everyday examples include senior citizen discounts at movie theaters and discounts on airline tickets for Saturday-night stayovers.

price discrimination: a situation in which different groups of consumers are charged different prices for the same good.

Some price discrimination is less noticeable because it occurs in geographically separated markets. Charging different prices in foreign markets and domestic markets is common. For example, Japanese cameras are less expensive in the United States than in Japan. In contrast, the price of luxury German cars in the United States is frequently higher than in Germany.

Volume or quantity discounts are another form of price discrimination. Higher prices are sometimes charged to customers who buy smaller amounts of an item. For example, electric utility firms sometimes charge more per kilowatt-hour to customers who use only a little electricity.

READING THE NEWS ABOUT... Barriers to Entry in the Microprocessor Industry

What makes a monopoly a monopoly? Intel may be the world's largest chip maker, but other companies have competed in the microprocessor chip market ever since Advanced Micro Devices won a series of legal battles in the 1990s that allowed it to make chips with its own designs. Now AMD argues that Intel, which can no longer rely on patents and copyrights to protect its monopoly, is using illegal tactics to win back its monopoly on this market. AMD claims that Intel's use of marketing subsidies to win sales, heavy discounts, and threatened retaliation against Intel customers for using AMD products amount to unfair barriers to entry. What do you think?

AMD Files Broad Suit Against Intel

Antitrust Case Argues Steps Were Set to Keep Monopoly Over Global Chip Market

By DON CLARK
Staff Reporter of THE WALL STREET JOURNAL
June 28, 2005

Advanced Micro Devices Inc. has filed a broad antitrust suit against **Intel** Corp., accusing its giant rival of using illegal inducements and coercion to dissuade companies from buying AMD's computer chips.

The suit, filed late yesterday in federal court in Delaware, alleges that Intel has engaged in a "relentless" global campaign to maintain a monopoly over microprocessors that serve as the electronic brains in most computers. AMD alleges that Intel used improper subsidies to win sales, and threatened retaliation against firms for using or selling AMD products.

Tom Beerman, an Intel spokesman, said it hadn't seen the suit, and said only "we believe our sales practices are both fair and consistent with federal antitrust law," he said.

The case pits a struggling challenger against the world's largest chip maker, in what promises to be a protracted and pivotal fight. Although launched by a company rather than the government, the case could have large-scale repercussions in the industry. So far Intel has faced less antitrust scrutiny than its software partner, **Microsoft** Corp., which was the subject of a major antitrust prosecution by the Justice Department, but it matches Microsoft's influence in setting technology standards and grabbing a huge share of computer-industry profits.

AMD's 48-page complaint, which follows a recent antitrust ruling against Intel in Japan, may prompt a debate on Intel marketing subsidies that have become a mainstay for computer makers with thin profit margins. If a court rules against such practices, dominant companies could lose an important tool to command customer loyalty.

The case could illuminate confidential dealings between Intel and other industry players. AMD's complaint lists examples of what it characterizes as bribes, threats or intimidation by Intel involving 12 computer makers, nine distributors and 17 retailers. Customers cited include **International Business Machines** Corp., **Hewlett-Packard** Co., **Dell** Inc., **Sony** Corp., **Toshiba** Corp. and **Gateway** Inc.

Consumers with Different Price Elasticities of Demand

Why is there price discrimination? Figure 10.9 shows a diagram of a monopoly that gives one explanation. Suppose the good being sold is airline travel between two remote islands, and suppose there is only one airline between the two islands. The two graphs in Figure 10.9 represent demand curves with different elasticities. On the

The allegations are based largely on discussions between AMD and customers. To document Intel's alleged behavior, AMD plans to seek subpoenas to obtain private email from those companies, and risk alienating industry executives by asking them to testify on its behalf.

"They need to sustain their complaint by customer testimony," said Eleanor Fox, a professor at the New York University School of Law, who isn't involved in the case. "Customers may not be so friendly to the idea."

Hector Ruiz, AMD's chief executive, said it has consulted with many Intel customers and partners, whom he expects to help in the litigation. "To a person, they are going to be glad that we put this on the table, though they may not come out and say so," he said.

Intel, of Santa Clara, Calif., commands more than 80% of the unit sales and 90% of the revenues in the market for x86 microprocessors, named for a set of instructions that help define what software a chip uses. AMD, based in nearby Sunnyvale, estimates that its share of x86 unit sales peaked in 2001 at 20.8% but ebbed to 15.8% by 2004.

AMD originally made chips using designs from Intel. That arrangement collapsed in the mid-1980s, setting off a series of legal battles that were settled in 1995. The agreement gave AMD rights to make x86 chips with original designs.

AMD's complaint alleges that, once it began making headway with new products in 1999, Intel fought back with illegal tactics. One focus is a variety of rebate that AMD says Intel distributes at the end of the quarter to customers who buy nearly all their chips from the company.

Because Intel chips have become a standard feature for many personal-computer models, Intel wins more than half of many companies' chip purchases, the AMD complaint states. In competing for the remainder, Intel can offer 8% to 10% discounts over all the chips they sell a company—an advantage AMD can counter only with prohibitive price cuts in bidding for a minority chunk of the business, AMD said.

"The last thing AMD fears is price competition," says Thomas McCoy, AMD's vice president of legal affairs and chief administrative officer. But Intel's rebates amount to "dictating to the industry how many AMD-based computers it can sell," he said.

Related issues were raised in March by Japan's Fair Trade Commission, which alleged that Intel violated antitrust laws by using rebates and marketing funds to dissuade five Japanese PC makers from buying from AMD and Transmeta Corp. The deals required the firms to buy 90% or 100% of their chips from Intel, or to limit specific lines of PCs to Intel chips, the Japanese agency said.

Intel chose not to contest that ruling.

Ms. Fox, the antitrust expert, predicted the case will turn on whether Intel can show that its marketing practices benefit consumers, often a powerful argument in antitrust cases. But one court ruling could help AMD, she noted.

In 2003 the U.S. Court of Appeals for the Third Circuit sided with LePage's Inc. in an antitrust case against 3M Co. over private-label transparent tape. The judges ruled against 3M's "bundled rebates," which offered incentives to customers who purchased 3M product lines in addition to its tape. Charles Diamond, an attorney for AMD at O'Melveny & Myers LLP, said its allegations are very similar to LePage's, noting that the same appeals court could be the final arbiter of AMD's case.

The case could go to trial in late 2006, Mr. McCoy said.

left is the demand for vacation air travel. Vacationers are frequently more price sensitive than businesspeople. They can be more flexible with their time; they can take a boat rather than the plane; they can stay home and paint the house. Hence, for vacationers, the price elasticity of demand is high. Business travelers, however, do not have much choice. As shown on the graph to the right, they are less sensitive to price.

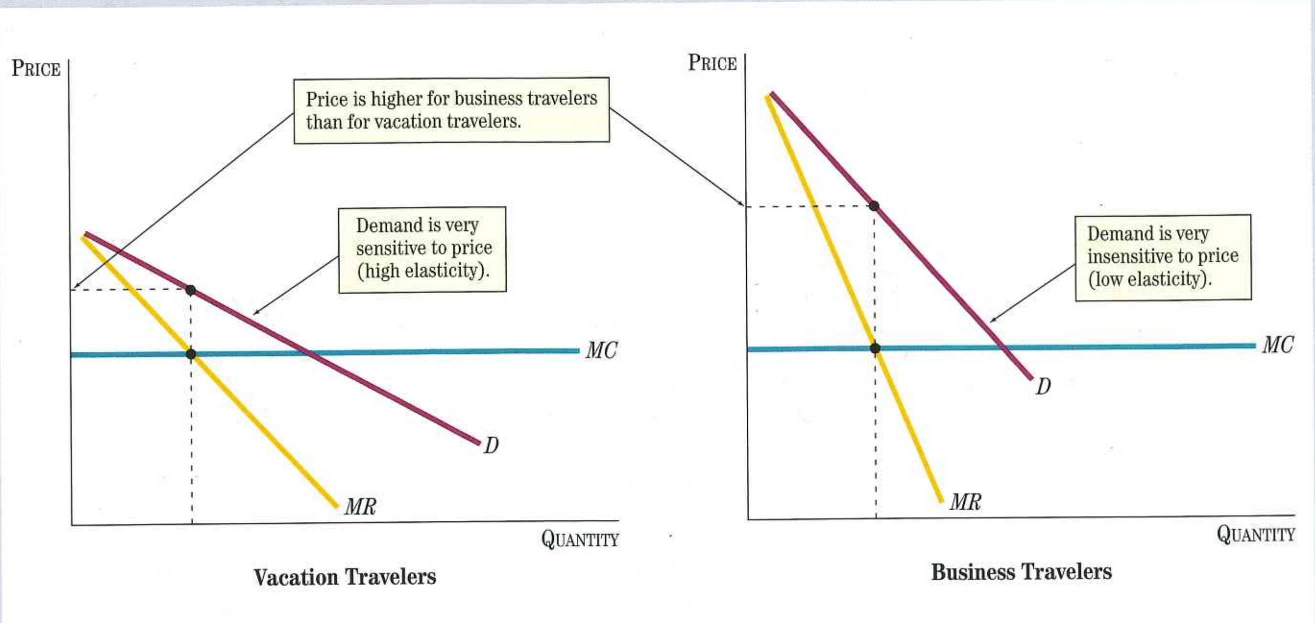


Figure 10.9 Price Discrimination Targeted at Different Groups

The monopolist has two groups of potential buyers for its travel services. For convenience, we assume the marginal cost curve is flat. The group on the left has a high price elasticity of demand. The group on the right has a low price elasticity of demand. If the monopolist can discriminate between the buyers, then it is optimal to charge a lower price to the high-elasticity group and a higher price to the low-elasticity group.

An important business meeting may require a businessperson to fly to the other island with little advance notice. For business travel, the price elasticity of demand is low. Difference between price elasticities is a key reason for price discrimination.

In Figure 10.9, notice that both groups have downward-sloping demand curves and downward-sloping marginal revenue curves. For simplicity, marginal cost is constant and is shown with a straight line.

Figure 10.9 predicts that business travelers will be charged a higher price than vacationers. Why? Marginal revenue equals marginal cost at a higher price for business travelers than for vacationers. The model of monopoly predicts that the firm will charge a higher price to those with a lower elasticity and a lower price to those with a higher elasticity.

In fact, this is the type of price discrimination we see with airlines. But how can the airlines distinguish a business traveler from a vacation traveler? Clothing will not work: A business traveler could easily change from a suit to an aloha shirt and shorts to get the low fare. One device used by some airlines is the Saturday-night stayover. Business travelers prefer to work and travel during the week. They value being home with family or friends on a Saturday night. Vacationers frequently do not mind extending their travel by a day or two to include a Saturday night, and they may want to vacation over the weekend. Hence, there is a strong correlation between vacation travelers and those who do not mind staying over a Saturday night. A good way to price-discriminate, therefore, is to charge a lower price to people who stay in their destination on a Saturday night and to charge a higher price to those who are unwilling to do so.

Price discrimination based on different price elasticity of demand requires that the firm be able to prevent people who buy at a lower price from selling the item to other people. Thus, price discrimination is much more common in services than in manufactured goods.

"Would it bother you to hear how little I paid for this flight?"

Quantity Discounts

Another important form of price discrimination involves setting prices according to how much is purchased. If a business makes 100 telephone calls a day, it probably has to pay a higher fee per call than if it makes 1,000 a day. Telephone monopolies can increase their profits by such a price scheme, as shown in Figure 10.10.

The single-price monopoly is shown in the bottom graph of Figure 10.10. Two ways in which the monopoly can make higher profits by charging different prices are shown in the top two panels. In both cases, there is no difference in the price elasticity of demand for different consumers. To make it easy, assume that all consumers are identical. The demand curve is the sum of the marginal benefits of all the consumers in the market.

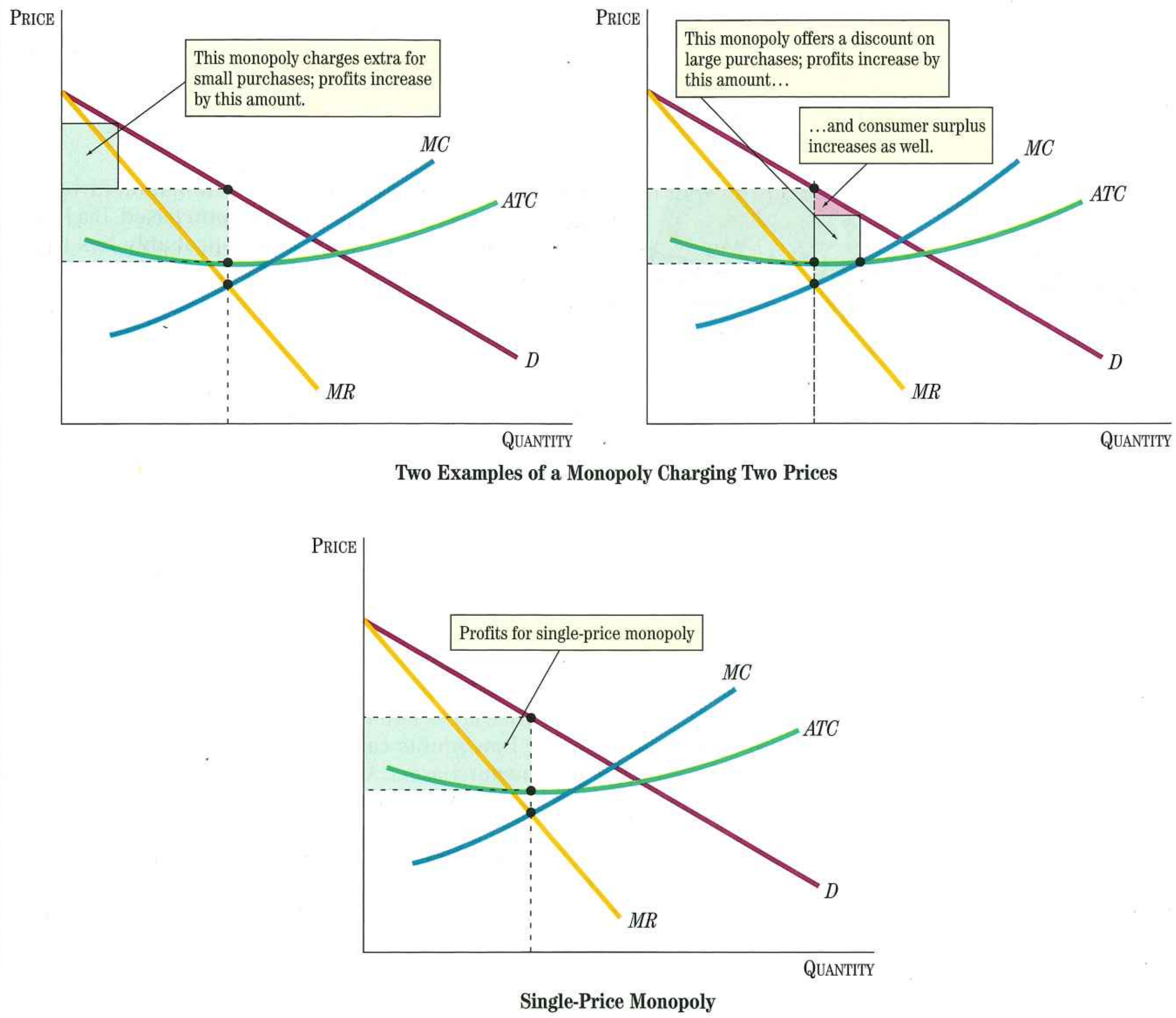
On the upper left, the firm sets a higher price for the first few items a consumer buys and a lower price for the remaining items. Frequent flier miles on airlines are an example of this kind of pricing. If you fly more than a certain number of miles, you get a free ticket. Thus, the per mile fare for 20,000 miles is less than the fare for 10,000 miles. As the diagram shows, profits for the firm are higher in such a situation. In the example at the left, the higher price is the fare without the discount.

On the upper right, we see how profits can be increased if the firm gives even deeper discounts to high-volume purchasers. As long as the high-volume purchasers cannot sell the product to the low-volume purchasers, there are extra profits to be made.

The upper right graph in Figure 10.10 illustrates an important benefit of price discrimination: It can reduce deadweight loss. With price discrimination, a monopoly actually produces more. For example, those who get a lower price because of frequent flier discounts may actually end up buying more. The result is that the airline has more flights. As already noted, the deadweight loss from a monopolist occurs because production is too low. If price discrimination allows more production, then it reduces deadweight loss.

Monopolies versus Other Firms with Market Power

The preceding examples of price discrimination involved monopolies, but price discrimination also occurs in firms that are not monopolies. United, American, and Delta all offer different fares to customers based on different price elasticities when they fly the same routes. We will see in Chapter 11 that firms can have some monopoly power—face downward-sloping demand as shown in Figures 10.9 and 10.10—even if they are not monopolies. For example, firms in industries in which one firm's products are slightly different from other firms' products have market power. The preceding explanation of price discrimination can, therefore, apply to such firms.



Two Examples of a Monopoly Charging Two Prices

Single-Price Monopoly

Figure 10.10
Price Discrimination Through Quantity Discounts or Premiums

The standard single-price monopoly is shown at the bottom. If the monopoly can charge a higher price to customers who buy only a little, profits can increase, as shown on the upper left. If the monopoly can give a discount to people who purchase a lot, it can also increase profits, as shown on the upper right. In this case, production increases.

REVIEW

- Because a monopolist has market power, it can charge different prices to different consumers as long as it can prevent the consumers from reselling the good.
- Price discrimination explains telephone pricing as well as the complicated airfares on airlines.
- Deadweight loss is reduced by price discrimination.

Conclusion

The model of a monopoly that we developed in this chapter centers on a key diagram, Figure 10.6 on page 267. Learning how to work with this diagram of a monopoly is very important. In fact, economists use this same diagram to describe any firm that has some market power, not just monopolies, as we show in Chapter 11. Before proceeding, it is a good idea to practice sketching this generic diagram of a monopoly and finding output, price, and profits for different positions of the curve.

From the point of view of economic efficiency, the economic performance of monopolies is not nearly as good as that of competitive industries. Output is too low, marginal benefits are not equal to marginal costs, and consumer surplus plus producer surplus is diminished. But when assessing these losses, the fact that the expectation of monopoly profits—even if temporary—is the inducement for firms to do research and develop new products must also be considered.

Nevertheless, the deadweight loss caused by monopolies provides a potential opportunity for government to intervene in the economy. In fact, the U.S. government actively intervenes in the economy either to prevent monopolies from forming or to regulate monopolies when it is not appropriate to break them apart. We look further into government prevention or regulation of monopolies in Chapter 12.

KEY POINTS

1. A monopoly occurs when only one firm sells a product for which there are no close substitutes. The world diamond market is nearly a monopoly. Many local markets for water, sewage, electricity, and telephones are monopolies.
2. A monopolist possesses market power in the sense that it can lower the market price by producing more or raise the market price by producing less.
3. The model of a monopoly assumes that the monopoly tries to maximize profits and that it faces a downward-sloping demand curve.
4. A monopoly's total revenue increases and then decreases as it increases production.
5. A monopoly chooses a quantity such that marginal revenue equals marginal cost.
6. A monopoly produces a smaller quantity and charges a higher price than a competitive industry; the lower production causes a deadweight loss.
7. It is frequently not possible to create a competitive industry out of a monopoly, in which case the comparison between monopoly and competition is hypothetical.
8. Monopolies exist because of economies of scale that make the minimum efficient size of the firm larger than the market, and because of barriers to entry, including government patents and licenses.
9. Many monopolies are short-lived; technological change can rapidly change a firm from a monopoly to a competitive firm, as exemplified by the long-distance telephone market.
10. A price-discriminating monopoly charges different prices to different customers depending on how elastic their demand is.

KEY TERMS

monopoly
barriers to entry
market power
price-maker
average revenue
price-cost margin
natural monopoly
contestable market
price discrimination

QUESTIONS FOR REVIEW

1. What is a monopoly?
2. What market power does a monopoly have?
3. How does a monopoly choose its profit-maximizing output and price?
4. Why does marginal revenue decline as more is produced by a monopoly?
5. Why is the marginal revenue curve below the demand curve for a monopoly but not for a competitive firm?
6. Why does a monopolist produce less than a competitive industry?

7. What forces tend to cause monopolies?
8. What is the deadweight loss from a monopoly?
9. What is price discrimination?
10. How does price discrimination reduce deadweight loss?

PROBLEMS

1. Suppose the price elasticity of demand for a drug is 1.25.
   a. Suppose only one monopolist firm produces the drug. If the monopolist cuts production by 15 percent, by what percentage does the price rise?
   b. Now suppose that the market is competitive, with 100 firms each supplying 1 percent of the market. If one of the competitive firms cuts its own production by 15 percent and the other 99 firms do not change production, by what percentage does the price rise?
   c. Will this decision by this one firm have any effect on the other firms in the industry? Explain.
2. The following table gives the total cost and total revenue schedule for a monopolist.

Quantity	Total Cost (in dollars)	Total Revenue (in dollars)
0	144	0
1	160	90
2	170	160
3	194	210
4	222	240
5	260	250
6	315	240
7	375	210

   a. Calculate the marginal revenue and marginal cost, and sketch the demand curve.
   b. Determine the profit-maximizing price and quantity, and calculate the resulting profit.
3. The following table gives the round-trip airfares from Los Angeles to New York offered by United Airlines.

Price	Advance Purchase	Minimum Stay	Cancellation Penalty
$ 418	14 days	Overnight on Saturday	100%
$ 683	3 days	Overnight on Saturday	100%
$1,900	None required	None required	None

Explain why United might want to charge different prices for the same route. Why are there minimum-stay requirements and cancellation penalties?

4. Sketch the diagram for a monopoly with an upward-sloping marginal cost curve that is earning economic profits. Suppose the government imposes a tax on each item the monopoly sells. Draw the diagram corresponding to this situation. How does this tax affect the monopoly's production and price? Show what happens to the area of deadweight loss.
5. Children, students, and senior citizens frequently are eligible for discounted tickets to movies. Is this an example of price discrimination? Explain the conditions necessary for price discrimination to occur and draw the graphs to describe this situation.
6. Why is it that firms need market power in order to price-discriminate? What other circumstances are required in order for a firm to price-discriminate? Give an example of a firm or industry that price-discriminates and explain how it is possible in that case.
7. Fill in the missing data on a monopolist in the following table:

Quantity of Output	Price	Total Revenue	Marginal Revenue	Marginal Cost	Average Total Cost
1	11				18.00
2	10				11.00
3	9				7.67
4	8				6.75
5	7				6.60
6	6				7.00
7	5				8.00

   a. At what quantity will the monopolist produce in order to maximize profits? What will be the price at this level of output? What will be the profits?
   b. What quantity maximizes total revenue? What is the elasticity of demand at that point? Why is this not the profit-maximizing quantity?
8. What is the price-cost margin for the typical competitive firm? What is the price-cost margin (at the profit-maximizing quantity) for the monopoly described in problem 7?
9. Calculate the deadweight loss of the monopoly described in the table on the facing page.

Problem 9

Quantity	Price	Total Revenue	Marginal Revenue	Total Cost	Marginal Cost	Profit
0	320	0	—	140	—	−140
2	305	610	305	158	9	452
4	290	1,160	275	168	5	992
6	275	1,650	245	188	10	1,462
8	260	2,080	215	228	20	1,852
10	245	2,450	185	296	34	2,154
12	230	2,760	155	392	48	2,368
14	215	3,010	125	522	65	2,488
16	200	3,200	95	702	90	2,498
18	185	3,330	65	962	130	2,368
20	170	3,400	35	1,312	175	2,088
22	155	3,410	5	1,762	225	1,648
24	140	3,360	−25	2,322	280	1,038

10. The first table below shows the marginal benefit schedule for the three buyers in a market. The second table below shows the marginal cost schedules for the three sellers in the market.

Quantity	*MB*—Linda	*MB*—Sue	*MB*—Pete
1	15	14	13
2	12	11	10
3	9	8	7
4	6	5	4
5	3	2	1

Quantity	*MC*—Max	*MC*—Scott	*MC*—Karen
1	3	2	1
2	6	5	4
3	9	8	7
4	12	11	10
5	15	14	13

Suppose the three sellers in the market merge to form a monopoly. The buyers continue to act independently. Assume that the marginal cost is the sum of the marginal costs of the three original sellers.

a. Compute the marginal revenue for the monopoly and plot it.

b. What output and what price do you predict the monopoly will choose?

c. What is the price-cost margin?

d. Show the loss of consumer surplus due to the monopoly.

e. Show the deadweight loss due to the monopoly.

11. Suppose you are an economic adviser to the president, and the president asks you to prepare an economic analysis of Monopoly, Inc., a firm that sells a patented device used in high-definition television sets. You have the following information about Monopoly, Inc.

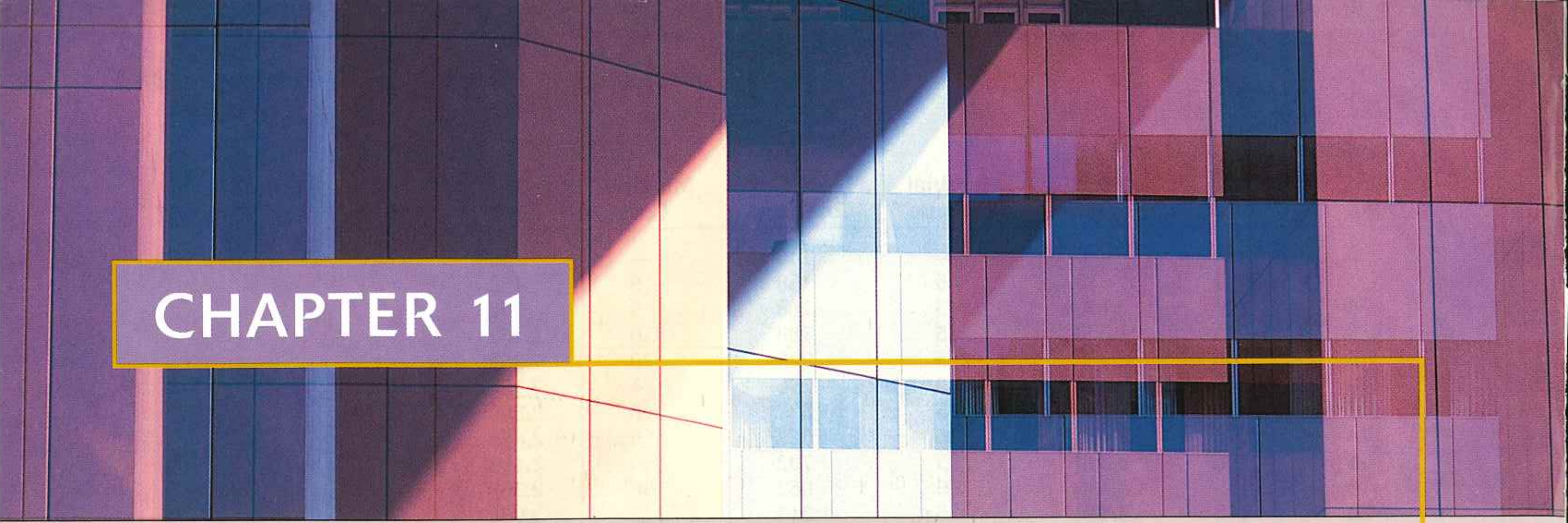
Quantity (millions)	Price	Marginal Cost
1	10	4
2	9	5
3	8	6
4	7	7
5	6	8
6	5	9
7	4	10
8	3	11
9	2	12
10	1	13

a. Given the data in the table, graphically show all the elements necessary to represent the monopolist's profit maximization. *Note:* You do not need to draw the average total cost curve.

b. What level of output does Monopoly, Inc., produce? What price does it sell this output at?

c. Does Monopoly, Inc., produce at the socially optimal level? Why or why not? Show any inefficiency on your graph.

d. Because of Monopoly, Inc.'s strong political lobby, the president is considering subsidizing the monopoly's production. As an economist, however, you are not concerned with politics, but want to ensure that this policy would not make the economy any worse off. Devise a subsidy whereby you could both satisfy the president's political needs and improve the efficiency of the economy. Why does your plan improve economic efficiency?



CHAPTER 11

Product Differentiation, Monopolistic Competition, and Oligopoly

When John Johnson launched his magazine business in 1942, he differentiated his products from existing products in a way that was valued by millions of African Americans. As a result, the new product lines, the magazines *Ebony* and *Jet,* were huge successes. Johnson became a multimillionaire, and his firm became the second largest black-owned firm in the United States. Similarly, when Liz Claiborne started her new clothing firm in 1976, she differentiated her products from existing products in a way that was valued by millions of American women. She offered stylish yet affordable clothes for working women, and she too was successful: By 1991, Liz Claiborne, Inc., was the largest producer of women's clothing in the world. Such stories of people finding ways to differentiate their products from existing products are told thousands of times a year, although not everyone is as successful as John Johnson and Liz Claiborne.

John Johnson's magazines and Liz Claiborne's suits and dresses were different in a way that was valued by consumers. Because their products were different from the products made by the many other firms in their industries—

magazine publishing and women's clothing, respectively—they each had market power in the sense that they could charge a higher price for their products and not lose all their customers. Thus, neither Johnson Publishing nor Liz Claiborne, Inc., was just another firm entering a competitive industry in which every firm sold the same product. But Johnson Publishing and Liz Claiborne, Inc., were not monopolies either; there were other firms in their industries, and they could not prevent entry into the industries by even more firms. As is typical of many firms, they seemed to be hybrids between a competitive firm and a monopoly.

In this chapter, we develop a model that is widely used by economists to explain the behavior of such firms. It is called *the model of monopolistic competition*. **Monopolistic competition** occurs in an industry with many firms and free entry, where the product of each firm is slightly differentiated from the product of every other firm. We contrast the predictions of this model with those of the models of competition and monopoly developed in previous chapters.

monopolistic competition: a market structure characterized by many firms selling differentiated products in an industry in which there is free entry and exit.

We also develop a *model of oligopoly* in this chapter, because product differentiation is not the only reason many firms seem to fall between the models of monopoly and competition. In an **oligopoly,** there are very few firms in the industry. Because there are very few firms, each firm has market power. The actions of any one firm can significantly affect the market price.

oligopoly: an industry characterized by few firms selling the same product with limited entry of other firms.

The firms' behavior is strategic in the sense that each needs to anticipate what the others will do and develop a strategy to respond. Such situations, whether in games or in industry, are very complex. Neither the model of a competitive industry, where no one firm can affect the price, nor the model of monopoly, where one firm completely dominates the market, adequately describes such a situation. Hence, there is a need for a model of oligopoly to explain situations in which a few firms produce goods and services and engage in strategic behavior: thinking about, anticipating, and reacting to the other firms' moves.

Figure 11.1 compares the models of monopolistic competition, oligopoly, monopoly, and competition. Over time, an industry can change from being a monopoly to monopolistic competition, to oligopoly, to competition, and back again, as a result of changes in the number of firms or the degree of product differentiation.

In order to emphasize the distinction between the models of competition and monopolistic competition or between the models of monopoly and monopolistic competition, the terms *pure competition* and *pure monopoly* are sometimes used. In this book, we simply use the terms *competition* and *monopoly*.

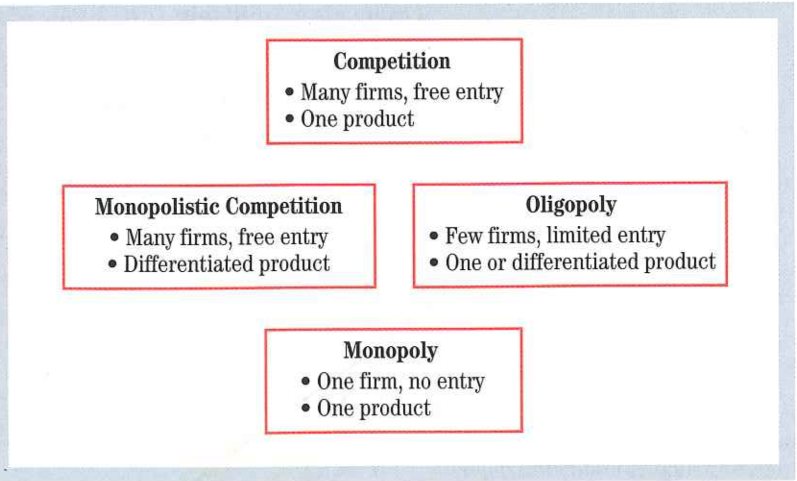


Figure 11.1
Four Types of Industries
Monopoly and competition are at the extreme ends. Monopolistic competition and oligopoly are in between.

Product Differentiation

product differentiation: the effort by firms to produce goods that are slightly different from other types of goods.

The effort by firms to fashion products that are different from other firms' products in ways that people value is called **product differentiation.** Product differentiation is pervasive in market economies. It leads to a great variety of consumer goods and capital goods. Goods for which there is no product differentiation, such as aluminum ingots or gold bullion, are called *homogeneous products,* meaning that they are all exactly the same.

Variety of Goods in a Market Economy

Product differentiation is obvious from a casual examination of the wide variety of goods in a modern market economy. Table 11.1 gives an indication of this wide vari-

Table 11.1
Variety: An Illustration of Product Differentiation

Item	Number of Different Types	Item	Number of Different Types
Automobile models	260	National soft drink brands	87
Automobile styles	1,212	Bottled water brands	50
SUV models	38	Milk types	19
SUV styles	192	Colgate toothpastes	17
Personal computer models	400	Mouthwashes	66
Movie releases	458	Dental flosses	64
Magazine titles	790	Over-the-counter pain relievers	141
New book titles	77,446	Levis jeans styles	70
Amusement parks	1,174	Running shoe styles	285
TV screen sizes	15	Women's hosiery styles	90
Frito-Lay chip varieties	78	Contact lens types	36
Breakfast cereals	340		

Source: 1998 Annual Report, Federal Reserve Bank of Dallas.

Product Differentiation versus Homogeneous Product
Even bottled water has become a highly differentiated product, more like breakfast cereal and soft drinks than like gold bullion, a homogeneous product for which there is no product differentiation.

ety. If you like to run, you have a choice of 285 different types of running shoes. You can choose among 340 different types of cereals for breakfast and wear 70 different types of Levis jeans.

The wide variety of products in a market economy contrasts starkly with the absence of such variety that existed in the once centrally planned economies of Eastern Europe and the Soviet Union. Stores in Moscow or Warsaw would typically have only one type of each product—one type of wrench, for example—produced according to the specifications of the central planners. There was even relatively little variety in food and clothing. One of the first results of market economic reform in these countries has been an increase in the variety of goods available.

Product differentiation is a major activity of both existing firms and potential new firms. Business schools teach managers that product differentiation ranks with cost cutting as one of the two basic ways in which a firm can improve its performance. An entrepreneur can enter an existing industry either by finding a cheaper way to produce an existing product or by introducing a product that is differentiated from existing products in a way that will appeal to consumers.

Product differentiation usually means something less than inventing an entirely new product. Aspirin was an entirely new product when it was invented; wrapping aspirin in a special coating to make it easier to swallow is product differentiation. Coke, when it was invented in 1886, was a new product, whereas Pepsi, RC Cola, Jolt Cola, Yes Cola, and Mr. Cola, which followed over the years, are differentiated products.

Product differentiation also exists for capital goods—the machines and equipment used by firms to produce their products. The large earthmoving equipment produced by Caterpillar is different from that produced by other firms, such as Komatsu of Japan. One difference is the extensive spare parts and repair service that go along with Caterpillar equipment. Bulldozers and road graders frequently break down and need quick repairs; by stationing parts distributorships and knowledgeable mechanics all over the world, Caterpillar can offer quick repairs in the event of costly breakdowns. In other words, the products are differentiated on the basis of service and a worldwide network.

Puzzles Explained by Product Differentiation

Product differentiation explains certain facts about a market economy that could be puzzling if all goods were homogeneous.

intraindustry trade: trade between countries in goods from the same or similar industries.

interindustry trade: trade between countries in goods from different industries.

Intraindustry Trade. Differentiated products lead to trade between countries of goods from the *same industry,* called **intraindustry trade.** Trade between countries of goods from *different industries,* called **interindustry trade,** can be explained by comparative advantage. Bananas are traded for wheat because one of these goods is grown better in warm climates and the other is grown better in cooler climates. But why should intraindustry trade take place? Why should the United States both buy beer from Canada and sell beer to Canada? Beer is produced in many different countries, but a beer company in one country will differentiate its beer from that of a beer company in another country. In order for people to benefit from the variety of beer, we might see beer produced in the United States (for example, Budweiser) being exported to Canada and, at the same time, see beer produced in Canada (for example, Molson) being exported to the United States. If all beer were exactly the same (a homogeneous commodity), such trade within the beer industry would make little sense, but it is easily understood when products are differentiated.

Advertising. Product differentiation also provides one explanation of why there appears to be so much advertising—the attempt by firms to tell consumers what is good about their products. If all products were homogeneous, then advertising would make little sense: A bar of gold bullion is a bar of gold bullion, no matter who sells it. But if a firm has a newly differentiated product in which it has invested millions of dollars, then it needs to advertise it to prospective customers. You can have the greatest product in the world, but it will not sell if no one knows about it. Advertising is a way to provide information to consumers about how products differ.

Economists have debated the role of advertising in the economy for many years. Many have worried about the waste associated with advertising. It is hard to see how catchy phrases like "It's the right one, baby" are providing useful information about Diet Pepsi to consumers. One explanation is that the purpose of the advertising in these cases is to get people to try the product. If they like it, they will buy more; if they do not like it, they will not—but without the ad they might not ever try it. Whatever the reason, advertising will not sell an inferior product—at least, not for long. For example, despite heavy advertising, Federal Express failed miserably with Zapmail—a product that guaranteed delivery of high-quality faxes of documents around the country within hours—because of the superiority of inexpensive fax machines that even small businesses could buy.

Others say that advertising is wasteful partly because it is used to create a *perception* of product differentiation rather than genuine differences between products. For example, suppose Coke and Pepsi are homogeneous products (to some people's tastes, they are identical). Then advertising simply has the purpose of creating a perception in people's minds that the products are different. If this is the case, product differentiation may be providing a false benefit, and the advertising used to promote it is a waste of people's time and effort.

Consumer Information Services. The existence of magazines such as *Consumer Reports* is explained by product differentiation. These magazines would be of little use to consumers if all products were alike.

Such services may also help consumers sort through exaggerated claims in advertising or help them get a better perception of what the real differences between products are. It is hard to sell an expensive product that ends up last on a consumer-rating list, even with the most creative advertising.

How Are Products Differentiated?

Altering a product's *physical characteristics*—the sharpness of the knife, the calorie content of the sports drink, the mix of cotton and polyester in the shirt, and so on—is the most common method of product differentiation. However, as the example of Caterpillar shows, products can be differentiated on features other than the physical characteristics. Related features such as low installation costs, fast delivery, large inventory, and money-back guarantees also serve to differentiate products.

Location is another important way in which products are differentiated. A Blockbuster Video or a McDonald's down the block is a very different product for you from a Blockbuster Video or a McDonald's 100 miles away. Yet only the location differentiates the product.

Time is yet another way to differentiate products. An airline service with only one daily departure from Chicago to Dallas is different from a service with 12 departures a day. Adding more flights of exactly the same type of air service is a way to differentiate the product. A 24-hour supermarket provides a different service from one that is open only during the day.

Convenience is increasingly being used by firms to differentiate products. How could peanut butter and jelly sandwiches, a standard for lunch, be more convenient? You can buy frozen peanut butter and jelly sandwiches on white bread. You can buy individually wrapped "slices" of peanut butter and jelly. How could a cup of coffee be more convenient? You could try coffee sold in a self-heating can that is hot exactly when you're ready to drink it.

The Optimal Amount of Product Differentiation at a Firm

Product differentiation is costly. Developing a new variety of spot remover that will remove mustard from wool (no existing product is any good at this) would require chemical research, marketing research, and sales effort. Opening another Lenscrafters (there are already hundreds in the United States) requires constructing a new store and equipping it with eyeglass equipment, trained personnel, and inventory.

But product differentiation can bring in additional revenue for a firm. The new spot remover will be valued by ice skaters and football fans who want to keep warm with woolen blankets or scarves but who also like mustard on their hot dogs. The people in the neighborhood where the new Lenscrafters opens will value it because they do not have to drive or walk as far.

The assumption of profit maximization implies that firms will undertake an activity if it increases profits. Thus, firms will attempt to differentiate their products if the additional revenue from product differentiation is greater than the additional costs. This is exactly the advice given to managers in business school courses. "Create the largest gap between buyer value . . . and the cost of uniqueness" is the way Harvard Business School professor Michael Porter puts it in his book *Competitive Advantage*.[1] If the additional revenue is greater than the additional cost, then business firms will undertake a product-differentiation activity.

[1] Michael Porter, *Competitive Advantage* (New York: Free Press, 1985), p. 153.

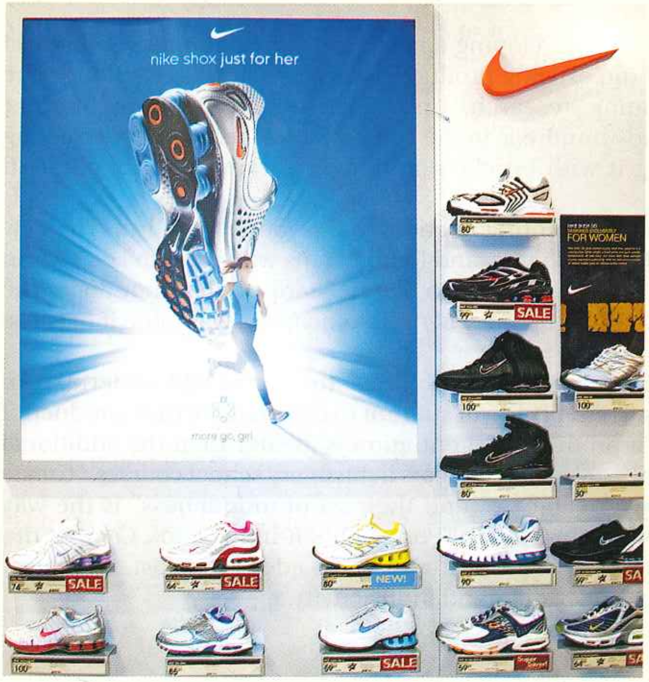
ECONOMICS IN ACTION

What's the Future of Product Differentiation?

How many types of running shoes do you think will be available for people to buy 10 years from now? There are now about 285 different types, but the number has grown tremendously in the last 25 years—there were only 5 types in the 1970s. This large increase in product differentiation is not unique to running shoes; it has occurred in virtually all markets. Colgate now produces 17 different types of toothpaste, compared with only 2 types in the 1970s. But will this rapid increase in product differentiation continue?

To determine whether an economic trend will continue, we first need to explain the trend. According to the theory of the optimal amount of product differentiation at a firm (see Figure 11.2 on page 291), a possible explanation for the increase in product differentiation is a reduction in its cost. Shifting the curve showing the "additional cost of product differentiation" down in Figure 11.2 would lead to more differentiated products. In fact, there is evidence that the cost of product differentiation has been reduced; computerized machines used to produce shoes make it easier to change the settings and alter the shape, thickness, or treads of rubber soles.

So the model explains the recent trends very well, and if the costs of product differentiation continue to fall in the future, we can expect a greater variety of products.



There is already evidence that computer technology is continuing to lower the cost of product differentiation. For example, a company called Footmaxx uses computers to determine a person's individual foot shape and gait characteristics. As the customer walks on a sensitive pad, the foot shape and pressure are captured many times throughout the gait cycle, and the data are fed into a computer, which prescribes a custom orthotic insole, designed to fit the foot exactly and correct the individual's gait. Nike's iD division has been letting sneakerheads design their own shoes online for several years; in May 2005, it went one step further by inviting sneaker fans to use their cell phones to customize a pair of shoes that was displayed on a 22-story screen in the middle of Times Square in New York City. After a minute-long session designing their shoe, the consumer could then download the design as wallpaper for his or her mobile phone or go online and buy the newly designed sneakers. The interactive experience combined both design and technology innovations. Other companies are following suit. Converse recently launched its own "Design Your Own" service on its web site.

In principle, it will be possible to choose a shoe that is unique to the individual—not only in style and color, but also in the shape of the foot and the characteristics of the gait. One can imagine more than a thousand types of running shoes—perhaps millions, one for every runner! Similar ideas are being developed for clothing, where a person's body is scanned by a laser and a shirt comes out exactly in the person's size.

Of course, these projections for the future require the *ceteris paribus* assumption that other things will remain the same. How important do you think that assumption is in this case? In particular, do you think consumers might change their behavior in response to such an explosion of product types?

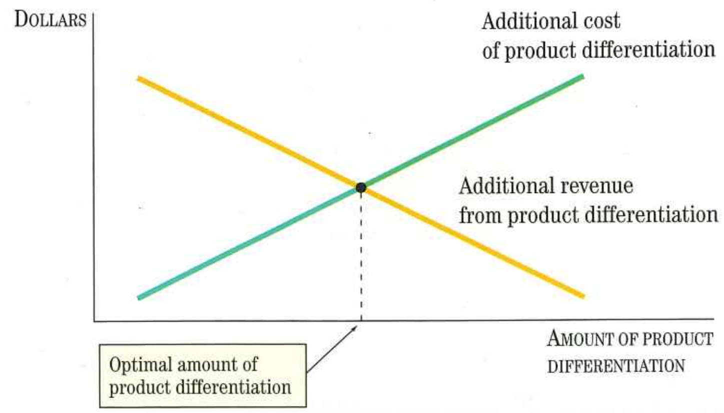


Figure 11.2
A Firm's Decision about Product Differentiation
Determining how much product differentiation a firm should undertake is a matter of equating the additional revenue from and additional cost of another differentiated product. (Note that these "additional cost" and "additional revenue" curves are analogous to marginal cost and marginal revenue curves except that they depend on the amount of product *differentiation* rather than the *quantity* of a particular product.)

For a given firm, therefore, there is an *optimal* amount of product differentiation that balances out the additional revenue and the additional cost of the product differentiation. This is illustrated in Figure 11.2, which shows the amount of product differentiation chosen by a firm. For a company that owns and operates a haunted house, the horizontal axis is the amount of gore and scary features in the haunted house. The additional revenue from adding more gore and scary features to a haunted house is shown by the downward-sloping line. While more gore and scary features attract additional customers, the additional revenue from increasing the amount of gore and scary features declines because there are only so many people who would consider visiting a haunted house in a given area. It is therefore increasingly difficult to attract additional customers. The additional cost of adding more gore and scary features to a haunted house is shown by the upward-sloping line. This additional cost increases because the cheapest effects that could be included for differentiation would be added first. The optimal amount of gore and scary features for a haunted-house operator is at the point where the additional revenue from more gore and scary features is just equal to the additional cost. Beyond that point, more gore and scary features would reduce profits, since the additional cost would exceed the additional revenue.

This is far from a trivial analysis for haunted-house owners. Theme parks are increasingly interested in attracting Halloween traffic, and more gore and scary features attract more customers. In some theme parks, Halloween is the largest event all year.

Using this analysis in practice is difficult because the revenue gains from product differentiation depend on what other firms do. The amount of additional revenue generated by additional gore and scary features in a haunted house depends on how much gore and how many scary features are included in other nearby haunted houses.

REVIEW

- Product differentiation is evident in the variety of products we see every day, in the absence of such variety in centrally planned economies, and in the proliferation of such variety after market reforms.
- Intraindustry trade, advertising, and consumer information are some of the facts that can be better explained by product differentiation.
- Products can be differentiated by physical characteristics, location, time, and convenience, among other features.
- Profit-maximizing firms will fashion a differentiated product if the additional revenue from doing so is greater than the additional cost.

Monopolistic Competition

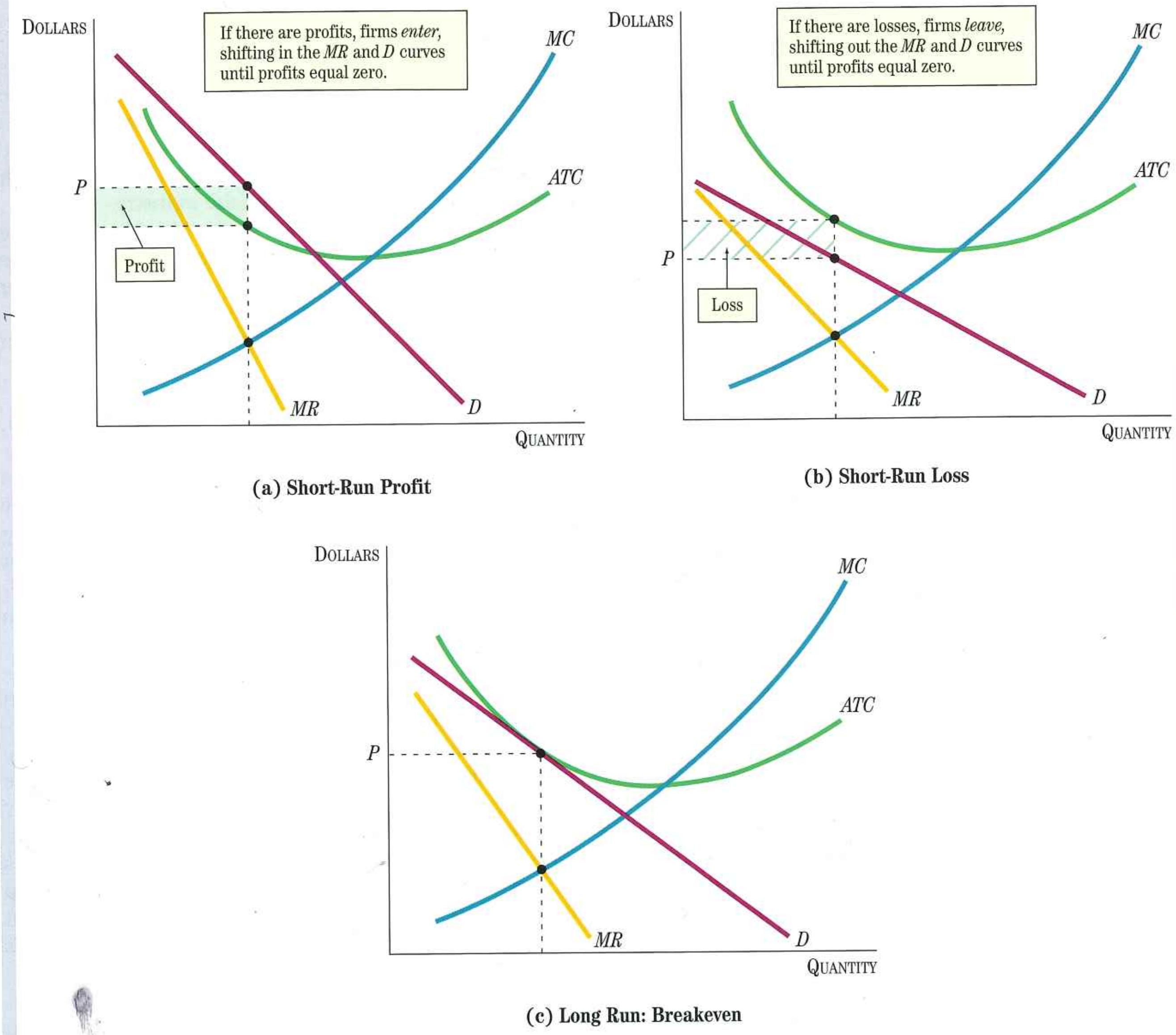
The model of monopolistic competition, first developed by Edward Chamberlin of Harvard University in the 1930s, is designed to describe the behavior of firms operating in differentiated product markets. Monopolistic competition gets its name from the fact that it is a hybrid of monopoly and competition. Recall that monopoly has one seller facing a downward-sloping market demand curve with barriers to the entry of other firms. Competition has many sellers, each facing a horizontal demand curve with no barriers to entry and exit. Monopolistic competition, like competition, has many firms with free entry and exit, but, as in monopoly, each firm faces a downward-sloping demand curve for its product.

The monopolistically competitive firm's demand curve slopes downward because of product differentiation. When a monopolistically competitive firm raises its price, the quantity demanded of its product goes down but does not plummet to zero, as in the case of a competitive firm. For example, if Nike raises the price of its running shoes, it will lose some sales to Reebok, but it will still sell a considerable number of running shoes because some people prefer Nike shoes to other brands. Nike running shoes and Reebok running shoes are differentiated products to many consumers. On the other hand, a competitive firm selling a product like wheat, which is a much more homogeneous product, can expect to lose virtually all its customers to another firm if it raises its price above the market price.

As we will see, free entry and exit is an important property of monopolistic competition. Because of it, firms can come into the market if there is a profit to be made or leave the market if they are running losses.

A Typical Monopolistic Competitor

Figure 11.3 illustrates the key features of the model of monopolistic competition. Each graph in Figure 11.3 shows a typical monopolistically competitive firm. At first glance, the graphs look exactly like the graph for a monopoly, introduced in Chapter 10. They should, because both monopolistic and monopolistically competitive firms face downward-sloping demand curves. However, the demand curve facing a monopolistically competitive firm has a different interpretation because there are other firms in the industry. The demand curve is not the market demand curve; rather, it is the demand curve that is *specific* to a particular firm. When new firms enter the industry—for example, when L.A. Gear enters with Nike and Reebok—the



(a) Short-Run Profit

(b) Short-Run Loss

(c) Long Run: Breakeven

Figure 11.3
Monopolistic Competition

Each graph shows a typical firm in a monopolistically competitive industry. Firms enter the industry if there are profits, as in graph (a). This will shift the demand and marginal revenue curves to the left for the typical firm because some buyers will switch to the new firms. Firms leave if there are losses, as in graph (b). This will shift the demand and marginal revenue curves to the right because the firms that stay in the industry get more buyers. In the long run, profits are driven to zero, as in graph (c).

demand curves specific to both Nike and Reebok shift to the left. When firms leave, the demand curves of the remaining firms shift to the right. The reason is that new firms take some of the quantity demanded away from existing firms, and when some firms exit, there is a greater quantity demanded for the remaining firms.

The difference between the graphs for a monopolist and a monopolistic competitor shows up when we move from the short run to the long run; that is, when firms enter and exit. This is illustrated in Figure 11.3. Note that the three graphs in the figure have exactly the same average total cost curve. The graphs differ from one another in that the location of the demand and marginal revenue curves relative to the average total cost curve is different in each. Graphs (a) and (b) represent the short run. Graph (c) represents the long run, after the entry and exit of firms in the industry.

Observe that the demand curve in graph (c) is drawn so that it just touches the average total cost curve. At this point, the profit-maximizing price equals average total cost. Thus, total revenue is equal to total costs, and profits are zero. On the other hand, in graphs (a) and (b), the demand curve is drawn so that there is either a positive profit or a negative profit (loss) because price is either greater than or less than average total cost.

▩ **The Short Run: Just Like a Monopoly.** Consider the short-run situation, before firms either enter or exit the industry. The monopolistic competitor's profit-maximization decision is like that of the monopoly. To maximize profits, it sets its quantity where marginal revenue equals marginal cost. Because the monopolistically competitive firm faces a downward-sloping demand curve, its profit-maximizing price and quantity balance the increased revenue from a higher price with the lost customers brought on by the higher price. The marginal-revenue-equals-marginal-cost condition achieves this balance. The profit-maximizing quantity of production is shown by the dashed vertical lines in graphs (a) and (b) of Figure 11.3.

For example, ForEyes, Lenscrafters, and Pearle Vision are monopolistic competitors in many shopping areas in the United States. Each local eyeglass store has an optometrist, but each offers slightly different services. At a shopping area with several of these eyeglass stores, if one of them raises prices slightly, then fewer people will purchase glasses there. Some people will walk all the way to the other end of the mall to the store with the lower-priced glasses. Others, however, will be happy to stay with the store that raised its prices because they like the service and the location. These outlets are not monopolists, but the downward slope of their demand curves makes their pricing decision much like that of monopolists. The slope of the demand curve for a monopolistic competitor may be different from that for a monopolist, but the qualitative relationship between demand, revenue, and costs—and the firm's decisions in setting quantity and price—is the same.

▩ **Entry and Exit: Just Like Competition.** Now consider entry and exit, which can take place over time. In the model of long-run competitive equilibrium in Chapter 9, we showed that if there were economic profits to be made, new firms would enter the industry. If firms were running losses, then firms would exit the industry. Only when economic profits were zero would the industry be in long-run equilibrium, with no tendency for firms either to enter or to exit.

In monopolistic competition, the entry and exit decisions are driven by the same considerations. If profits are positive, as in graph (a) of Figure 11.3, firms have incentive to enter the industry. Consider the market for hair products. Suave products are similar in appearance and function to other more expensive brands. If another producer has a top-selling shampoo, Suave will enter the market, selling a similar shampoo. If profits are negative, as shown in graph (b) of Figure 11.3, firms have incentive to exit the industry. Caribou Coffee, formerly located near the University of Michigan in Ann Arbor, closed because its costs exceeded its revenue. Demand for coffee was not high enough to support the number of coffee shops near campus.

As we move from the short run to the long run, the demand curve for each of the old firms will tend to shift to the point at which the demand curve and the average total cost curve are tangent—that is, the point where the two curves just touch and have the same slope. Entry into the industry will shift the demand curve of each existing firm to the left, and exit will shift the demand curve of each remaining firm to the right.

We now know why the demand and marginal revenue curves shift in this way. With new firms entering the market, the existing firms will be sharing their sales with the new firms. If Suave sells a new brand of shampoo similar to Pantene shampoo, then some consumers who had been buying Pantene will instead buy Suave's similar new shampoo. The demand for Pantene and other shampoos will therefore decline due to the availability of Suave's new product. Thus, the existing firms will see their demand curves shift to the left—each one will find it sells less at each price. The differences in the positions of the demand (and marginal revenue) curves in the short run and long run illustrate this shift. This shift in the demand curve occurs because new firms in the market are taking some of the demand, not because consumers have shifted their tastes away from the product. The shift in the demand curve causes each firm's profits to decline, and eventually profits decline to zero. (Recall that these are economic profits, not accounting profits, and are therefore a good measure of the incentive for firms to enter the industry.)

The case of negative profits and exit is similar. If demand is such that firms are running a loss, then some firms will exit the industry, causing the demand curve facing the remaining firms to shift to the right, until the losses (negative economic profits) are driven to zero. When Caribou Coffee closed in Ann Arbor, University of Michigan students bought coffee at other nearby coffee shops instead, increasing the demand for coffee at these nearby shops. This is illustrated by comparing graph (b) of Figure 11.3, where there are losses in the short run, with graph (c), where there are zero profits.

The Long-Run Monopolistically Competitive Equilibrium

There are two differences between monopolistically competitive firms and competitive firms in the long run. To see these differences, consider Figure 11.4, which replicates graph (c) of Figure 11.3, showing the position of the typical monopolistic competitor in long-run equilibrium, after entry and exit have taken place.

First, observe that price is greater than marginal cost for a monopolistically competitive firm. This was also true for the monopoly; it means that the market is not as efficient as a competitive market. Production is too low because the marginal benefit of additional production is greater than the marginal cost. Because each firm has some market power, it restricts output slightly and gets a higher price. The sum of producer plus consumer surplus is reduced relative to that in a competitive market. In other words, there is a loss of efficiency—a deadweight loss.

Second, as shown in Figure 11.4, the quantity produced is not at the minimum point on the average total cost curve, as it was for the competitive industry. That is, the quantity that the monopolistic competitor produces is at a higher-cost point than the quantity the perfect competitor would produce. Thus, monopolistically competitive firms operate in a situation of **excess costs.** If each firm expanded production and lowered its price, average total cost would decline. Each firm operates with some **excess capacity** in the sense that it could increase output and reduce average total cost. The firms choose not to do so because they have some market power to keep their prices a little higher and their output a little lower than that. Their market power comes from the downward-sloping demand curve they face. For

excess costs: costs of production that are higher than the minimum average total cost.

excess capacity: a situation in which a firm produces below the level that gives the minimum average total cost.

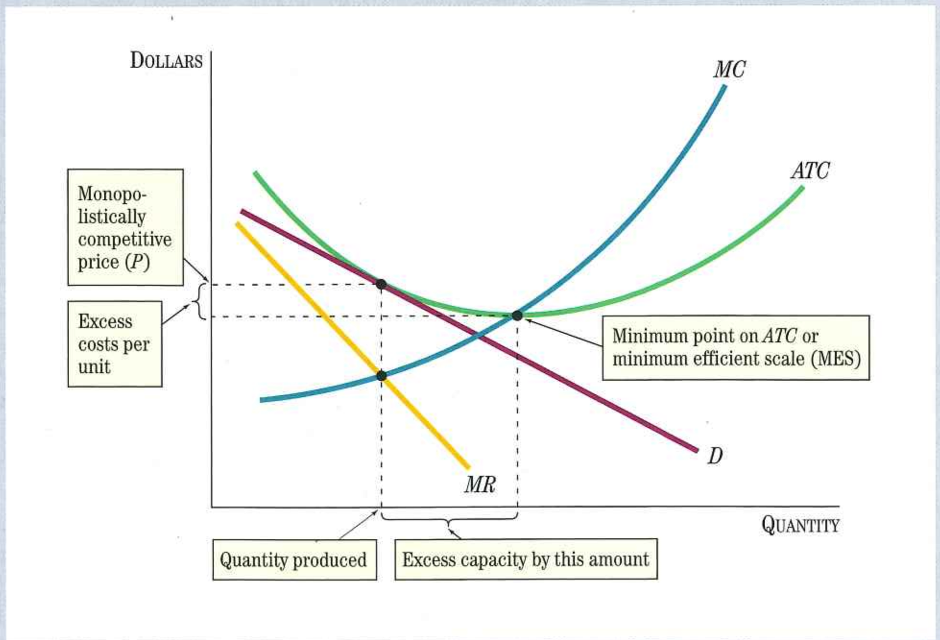


Figure 11.4
Excess Costs per Unit and Excess Capacity with Monopolistic Competition
In the long run, profits are zero for a monopolistically competitive firm, but the firm does not produce the quantity that minimizes average total cost. If the firm increases production, costs per unit will decline. In this sense, the firm operates at less than full capacity; it has excess capacity.

example, each coffee shop charges a little more and sells slightly fewer cups of coffee than it would in a perfectly competitive market.

Comparing Monopoly, Competition, and Monopolistic Competition. Table 11.2 compares the different effects of competition, monopoly, and monopolistic competition.

A competitive firm will produce the quantity that equates price and marginal cost. A competitive market is efficient in that consumer surplus plus producer surplus is maximized and there is no deadweight loss. Average total cost is minimized.

In monopoly, price is greater than marginal cost. A monopoly is inefficient because consumer surplus plus producer surplus is not maximized, so there is deadweight loss. Moreover, average total cost is not minimized. Economic profits remain positive because firms cannot enter the market.

Table 11.2

Type of Model	Price	Deadweight Loss?	Average Total Cost Minimized?	Profit in Long Run?
Competition	$P = MC$	No	Yes	No
Monopolistic competition	$P > MC$	Yes	No	No
Monopoly	$P > MC$	Yes	No	Yes

In monopolistic competition, price is also greater than marginal cost. Thus, consumer surplus plus producer surplus is not maximized and there is deadweight loss; average total cost is not minimized. However, profits are zero in the long-run equilibrium because of entry and exit. Monopolistic competition does not result in as efficient an outcome as competition. Monopolistic competition, as well as monopoly, is inefficient.

Product Variety versus Deadweight Loss. When comparing monopolistic competition with competition, we must recognize—as with the comparison of monopoly and competition in the last chapter—that replacing monopolistic competition with competition may be an impossibility or require a loss to society. Remember that product differentiation is the key reason for monopolistic competition. We showed in the previous section that the variety of products that comes from product differentiation is usually something that consumers value. Some people like having both Pepsi and Coke. Roads and airports are better because of the different capabilities of earthmoving equipment sold by Caterpillar and Komatsu. Thus, eliminating monopolistic competition by having a single competitive product, whether Coksi or Catematsu, even if it were possible, would probably reduce consumer surplus by more than the gain that would come from competition over monopolistic competition.

More generally, product differentiation may be of sufficient value to consumers that it makes sense to have monopolistically competitive firms despite the deadweight loss. Or, to state it somewhat differently, the deadweight loss from monopolistic competition is part of the price consumers pay for the variety or the diversity of products.

REVIEW

- The model of monopolistic competition is a hybrid of competition and monopoly. Entry and exit are possible, as in competition, but firms see a downward-sloping demand curve, as in monopoly, although there are many firms.
- The analysis of monopolistic competition in the short run is much like that of monopoly, but entry and exit lead to zero economic profits in the long run.
- Monopolistic competitors produce less than competitive firms and charge prices higher than marginal costs. Thus, there is a deadweight loss from monopolistic competition. In the long run, monopolistic competition produces less than the quantity that would minimize average total cost.
- The deadweight loss and excess costs can be viewed as the price of product variety.

Oligopoly

Thus far, we have seen two situations in which firms have market power: monopoly and monopolistic competition. But those are not the only two. When there are *very few* producers in an industry—a situation termed *oligopoly*—each firm can have an influence on the market price even if the goods are homogeneous. For example, if Saudi Arabia—one of the major producers of crude oil in the world and a member of the Organization of Petroleum Exporting Countries (OPEC)—decides to cut its production of crude oil, a relatively homogeneous commodity, it can have a significant effect on the world price of oil. However, the effect on the price will depend on what other producers do. If the other producing countries—Iran, Kuwait, and so on—increase their production to offset the Saudi cuts, then the price will not change by much. Thus, Saudi Arabia, either through formal discussion with other oil-producing countries in OPEC or by guessing, must take account of what the other producers will do.

Such situations are not unusual. The managers of a firm in an industry with only a few other firms know that their firm has market power. But they also know that the other firms in the industry have market power too. If the managers of a firm make the right assessment about how other firms will react to any course of action they take, then their firm will profit. This awareness and consideration of the market power and the reactions of other firms in the industry is called **strategic behavior.** Strategic behavior also may exist when there is product differentiation, as in monopolistically competitive industries, but to study and explain strategic behavior, it is simpler to focus on oligopolies producing homogeneous products.

strategic behavior: firm behavior that takes into account the market power and reactions of other firms in the industry.

A common approach to the study of strategic behavior of firms is the use of **game theory,** an area of applied mathematics that studies games of strategy like poker or chess. Game theory has many applications in economics and the other behavioral sciences. Because oligopoly behavior has many of the features of games of strategy, game theory provides a precise framework to better understand oligopolies.

game theory: a branch of applied mathematics with many uses in economics, including the analysis of the interaction of firms that take each other's actions into account.

An Overview of Game Theory

Game theory, like the basic economic theory of the firm and consumer (described in Chapters 5 and 6 of this book), makes the assumption that people make purposeful choices with limited resources. More precisely, game theory assumes that the players in a game try to maximize their payoffs—the amount they win or lose in the game. Depending on the application, a payoff might be measured by utility, if the player is a person, or by profits, if the player is a firm.

However, game theory endeavors to go beyond basic economic theory in that each player takes explicit account of the actions of each and every other player. It asks questions like: "What should Mary do if Deborah sees her and raises her by $10?" The aims of game theory are to analyze the choices facing each player and to design utility-maximizing actions, or strategies, that respond to every action of the other players.

An important example in game theory is the game called **prisoner's dilemma,** illustrated in Figure 11.5. The game is between Ann and Pete, two prisoners who have been arrested for a crime that they committed. The **payoff matrix** shown in Figure 11.5 has two rows and two columns. The two columns for Ann show her options, which are labeled at the top "confess" and "remain silent." The two rows for Pete show his options; these are also labeled "confess" and "remain silent." Inside the boxes, we see what happens to Ann and Pete for each option, confess or remain silent. The top right of each box shows what happens to Ann. The bottom left of each box shows what happens to Pete.

prisoner's dilemma: a game in which individual incentives lead to a nonoptimal (noncooperative) outcome. If the players can credibly commit to cooperate, then they achieve the best (cooperative) outcome.

payoff matrix: a table containing strategies and payoffs for two players in a game.

The police already have enough information to get a conviction for a lesser crime, for which Ann and Pete would each get a 3-year jail sentence. Thus, if both Ann and Pete remain silent, they are sent to jail for 3 years each, as shown in the lower right-hand corner of the table.

But Ann and Pete each have the option of confessing to the more serious crime that they committed. If Ann confesses and Pete does not, she gets a reward. If Pete confesses and Ann does not, he gets a reward. The reward is a reduced penalty: The jail sentence is only 1 year—not as severe as the 3 years it would be if the prosecutor

		Ann	
		Confess	Remain Silent
Pete	Confess	5 5	7 1
	Remain Silent	1 7	3 3

Figure 11.5
Two Prisoners Facing a Prisoner's Dilemma
Pete and Ann are in separate jail cells, held for a crime they *did* commit. The punishment for each—in years in jail—is given in the appropriate box and depends on whether they both confess or they both remain silent or one confesses while the other remains silent. The top right of each box shows Ann's punishment; the bottom left of each box shows Pete's punishment.

had no confession. However, the penalty for being convicted of the more serious crime in the absence of a confession is 7 years. Thus, if Ann confesses and Pete does not, he gets a 7-year sentence. If both confess, they each get a 5-year sentence.

What should Pete and Ann do? The answer depends on their judgment about what the other person will do. And this is the point of the example. Ann can either confess or remain silent. The consequences of her action depend on what Pete does. If Ann confesses and Pete confesses, she gets 5 years. If Ann confesses and Pete remains silent, Ann gets 1 year. If Ann remains silent and Pete remains silent, she gets 3 years. Finally, if Ann remains silent and Pete confesses, she gets 7 years. Pete is in the same situation that Ann is.

Think about a strategy for Ann. Ann is better off confessing, regardless of what Pete does. If Ann confesses and Pete confesses, Ann gets 5 years rather than 7 years. If Ann confesses and Pete remains silent, then Ann gets 1 year rather than 3 years. Hence, there is a great incentive for Ann to confess because she does better in either case.

Pete is in the same situation. He can compare what his sentence would be whether Ann confesses or remains silent. In this case, Pete is better off confessing regardless of whether Ann confesses or remains silent.

What this reasoning suggests is that both Ann and Pete will confess. If they both had remained silent, they would have gone to jail for only 3 years, but the apparently sensible strategy is to confess and go to jail for 5 years. This is the prisoner's dilemma. The case where both remain silent is called the **cooperative outcome** of the game because to achieve this, they would somehow have to agree in advance not to confess and then keep their word. The case where both confess is called the **noncooperative outcome** of the game because Pete and Ann follow an "everyone for himself or herself" strategy. Note that the cooperative outcome is preferred to the noncooperative outcome by both Pete and Ann, yet both choose the option that results in the noncooperative outcome.

The mathematician and Nobel laureate in economics John Nash defined the noncooperative equilibrium—which economists call a **Nash equilibrium**—as a set of strategies from which no player would like to deviate unilaterally—that is, no player would see an increase in his or her payoff by changing his or her strategy while the other players keep their strategies constant.

cooperative outcome: an equilibrium in a game where the players agree to cooperate.

noncooperative outcome: an equilibrium in a game where the players cannot agree to cooperate and instead follow their individual incentives.

Nash equilibrium: a set of strategies from which no player would like to deviate unilaterally.

CASE STUDY

Applying Game Theory

How do we apply game theory to the strategy of firms in an oligopoly? To make the application easier, focus on the case where there are only two firms. This is a particular type of oligopoly called *duopoly*. Let's first introduce an example of how to analyze a simple duopoly with game theory, and then we can generalize to more complex problems.

A Duopoly Game

The town of Pumpkinville announces that on October 10 it will hold a farmer's market where folks can buy giant pumpkins to carve in time for Halloween. Jack and Jill are the only two producers of giant pumpkins in Pumpkinville—Jack has a farm 5 miles east of town, while his competitor, Jill, has a farm 5 miles west of town. Back in April, Jack and Jill planted the seeds, and they cared for the pumpkins during the summer. Today is October 9. Jack and Jill each have 60 giant pumpkins ready to harvest, and they have to decide how many pumpkins they should harvest and transport to the market. The farmers are profit maximizers, and they take their costs and revenues into account. All the costs until today (seeds, fertilizer, water, labor, and so on) are sunk, cannot be altered, and should not affect the decision of whether to send the pumpkins to market or let them rot on the ground. The only relevant cost is the $1 per pumpkin for harvest and transportation.

Jack and Jill also know that the townsfolk love their pumpkins, but they are not willing to pay *any* price. The market demand for giant pumpkins is Price = $241 − $2 × Quantity, where the quantity is the sum of what Jack and Jill independently and simultaneously bring to the market on October 10. For example, if Jack decides to harvest and transport 60 pumpkins, while Jill decides to send only 30, then the total quantity supplied to the market will be 90, and the market price for giant pumpkins will be $61 (241 − 2 × 90). As you can see, Jack's decision will influence the price that Jill receives for her pumpkins, and vice versa. This situation is perfect for game theoretic analysis.

To simplify our analysis, let's assume that Jack's and Jill's strategies are limited to three actions: They can bring either 30, 40, or 60 pumpkins to the market. With the information on cost, prices, and available strategies, Jack and Jill can build the payoff matrix in Figure 11.6.

The first step is to build the skeleton of the matrix. We know there are two players—Jack and Jill—and three possible actions—30, 40, or 60 pumpkins. Thus, we build a three-by-three matrix with a total of nine blank boxes, one for each combination of Jack's and Jill's actions—the boxes are numbered for easy reference. Each of these boxes will be filled with the profits that Jack and Jill obtain given their actions. For example, in box 1 both Jack and Jill choose to harvest and transport only 30 pumpkins to the market. Let's calculate Jack's payoffs first. Total revenue will be the price of a pumpkin times the number of pumpkins sold by Jack. For this first box, Jack sells 30 pumpkins, while the market quantity is 60 pumpkins (30 from Jill's farm and 30 from Jack's), so the price is $121 (241 − 2 × 60) and total revenue is $3,630 ($121 times 30 pumpkins). Jack's relevant cost is $30 ($1 times 30 pumpkins). We subtract $30 from $3,630, and we get a payoff of $3,600 for Jack, which we write on the bottom left corner of the first box. The calculation for Jill is similar, and it also yields $3,600 (top right corner of the first box). The rest of the payoff matrix in Figure 11.6 is calculated in a similar way.

		Jill		
		30 pumpkins	40 pumpkins	60 pumpkins
Jack	30 pumpkins	1 $3,600 $3,600	2 $4,000 $3,000	3 $3,600 $1,800
	40 pumpkins	4 $3,000 $4,000	5 $3,200 $3,200	6 $2,400 $1,600
	60 pumpkins	7 $1,800 $3,600	8 $1,600 $2,400	9 $0 $0

Figure 11.6
Payoff Matrix for Jack and Jill
The payoff matrix contains the profits for Jack and Jill for every possible combination of their actions. For example, in box 6, Jack sends 40 pumpkins to the market, while Jill delivers 60; the market quantity will be 100, and the price will be $41 (241 − 2 × 100). Jack's revenue will be $1,640 (40 × $41) and his cost will be $40 ($1 per pumpkin), and thus his payoff is $1,600 ($1,640 − $40). Jill receives the same price per pumpkin, but since she sold 60, her revenue will be higher ($2,460), with costs of $60 and a payoff of $2,400.

Jack and Jill are aware of the nine possible outcomes, and the question is how each of them is going to choose a quantity to deliver to the farmer's market. Remember that today is October 9, and that the farmers have only one shot at this decision. Once they harvest their pumpkins and bring them to the market on October 10, it will be too late to change their choices. So Jack and Jill engage in a mental exercise, trying to figure out what each other will do.

Put yourself in Jill's shoes. She can easily see that her maximum payoff of $4,000 happens when she sends 40 pumpkins to the market while Jack sends only 30. However, Jill knows that Jack is a profit maximizer too, and if Jill sells 40 pumpkins, then Jack can increase his payoff from $3,000 to $3,200 by selling 40 instead of 30 pumpkins. So box 2 cannot be a solution to Jill's and Jack's problem, and neither can box 4. Jill may soon realize that box 1 provides the highest combined payoff ($7,200), and she may wonder whether that might be a feasible solution. However, if Jill chooses to sell 30 pumpkins, once again Jack will have the incentive to increase his production and sell 40 pumpkins, leaving Jill with a lower profit. Jill has the same incentive to sell 40 pumpkins, so box 1 does not work either.

At this point you can guess that we are looking for a combination of strategies from which neither player would like to deviate unilaterally—that is, a Nash equilibrium. In box 5, we find that neither Jack nor Jill wants to produce more or less pumpkins, given the production of their competitor. When Jack and Jill sell 40 pumpkins each, for a payoff of $3,200 each, we have the only Nash equilibrium of the game—check all remaining boxes yourself—and economics and game theory predict that that will be the outcome of the duopoly when the firms are choosing quantities.

Competition in Quantities versus Competition in Prices. In this example, Jack's and Jill's decision variable is the number of pumpkins. Price is left to be determined by the market demand, given the total number of pumpkins. This model of oligopoly is called *Cournot competition* in honor of the French economist Augustin Cournot, who created the original version of this model in 1838 (Cournot did not use game theory and the concept of Nash equilibrium, which was not invented until 1950).

Instead of competing in quantities, oligopolists can also compete in prices; this is called *Bertrand competition* for the French mathematician who reworked Cournot's model in terms of prices in 1883.

Comparison with Monopoly and Perfect Competition. Figure 11.7 shows the demand and marginal cost curves for pumpkins. The intersection of demand and marginal cost represents the competitive equilibrium, where Jack and Jill each supply 60 pumpkins at a price of $1 and obtain zero profits. The maximum combined payoff is the monopoly solution, which occurs when each farmer sells 30 pumpkins at a price of $121 per pumpkin. The Cournot solution lies between the monopoly and competitive equilibria in terms of price, quantity, and profit.

We do not analyze the Bertrand model in this book, but it may be interesting to note that it predicts that oligopolists will charge the same price and sell the same number of units as competitive firms would.

explicit collusion: open cooperation of firms to make mutually beneficial pricing or production decisions.

Collusion. The firms know that their combined profits can be maximized if they act together as a monopolist. There are three ways in which firms might act together. The first is by **explicit collusion,** in which the managers communicate with

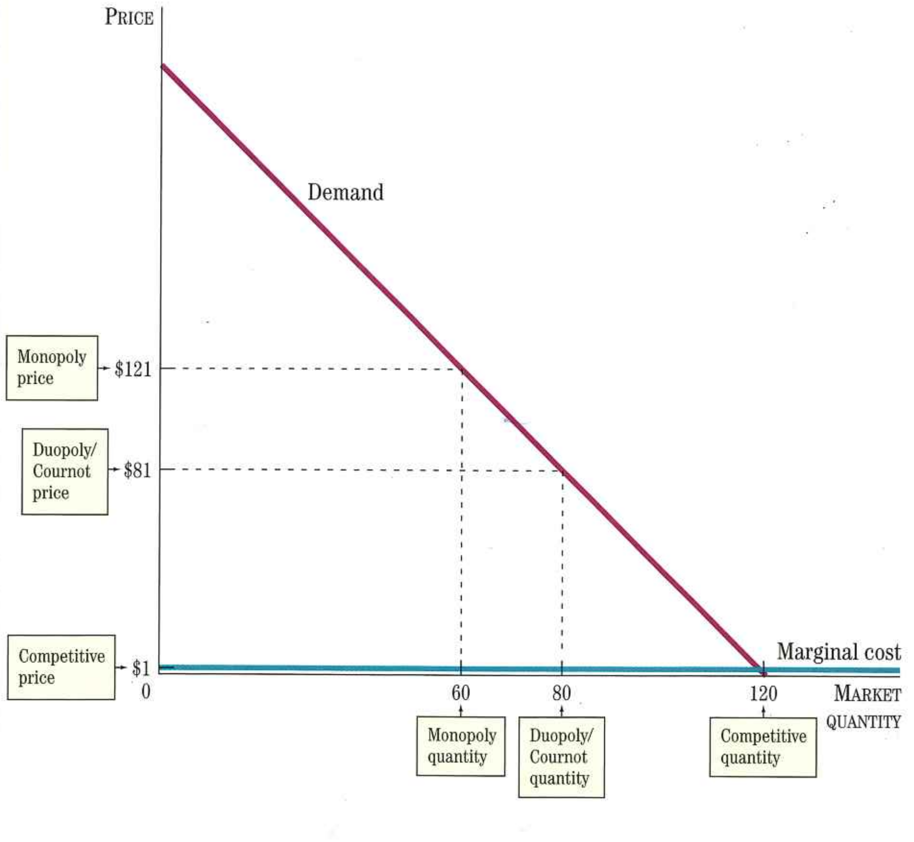


Figure 11.7
Comparison of Monopoly, Duopoly, and Competitive Equilibria
Prices and quantities for a Cournot duopoly lie between the equilibria for a monopoly and a competitive market.

each other and agree to fix prices or cut back on production. Although explicit collusion is illegal, it still happens. In the 1980s and 1990s, several firms in Florida and Texas were found guilty of agreeing to fix prices for milk sold to schools. In 1998, the firm Ucar International was found guilty of conspiring with other firms to fix prices and squelch competition in the market for graphite electrodes, a component of steelmaking furnaces. The governments of many countries that produce oil routinely collude to cut back production and raise prices. A group of producers that coordinates its pricing and production decisions is called a **cartel.**

cartel: a group of producers in the same industry who coordinate pricing and production decisions.

tacit collusion: implicit or unstated cooperation of firms to make mutually beneficial pricing or production decisions.

price leader: the price-setting firm in a collusive industry in which other firms follow the leader.

Second, there might be **tacit collusion,** where there is no explicit communication between firms, but firms keep prices high by regularly following the behavior of one firm in the industry. The dominant firm is sometimes called a **price leader.**

Third, the firms could merge, but that also might be illegal in the United States, as we discuss in Chapter 12.

Incentives to Defect. Notice that the duopoly situation has some similarities to the prisoner's dilemma. Just as there is an incentive for Pete and Ann to confess, there is an incentive for Jill and Jack to deviate from producing 30 pumpkins each, the solution that would give them the maximum combined payoff, as if they were behaving as one monopolist. In oligopoly, game theory predicts that unless there is a way to bind each firm to cooperation, there is a tendency to defect. Since the defection results in a lower than monopoly price, consumers gain from the defection, and deadweight loss is reduced.

Incentives to Cooperate: Repeated Games

Although the prisoner's dilemma and the Cournot duopoly suggest that there is a tendency to the noncooperative outcome, there is a difference between the situation of the prisoners Ann and Pete and the farmers Jack and Jill. Firms will presumably have future opportunities to interact. Pumpkinville's farmer's market will probably be open next year and for many years. If the same game is to be played year after year—a repeated game—then the firms might be able to build up a reputation for not defecting.

Experimental economists have conducted experiments in which two people play the same prisoner's dilemma game over and over again. (The people in the experiments are given small monetary rewards rather than jail penalties!) These experiments indicate that people frequently end up using strategies that lead to a cooperative outcome. A typical strategy people use is called "tit-for-tat." Using a tit-for-tat strategy, one player regularly matches, *in the next game,* the actions of the other player *in the current game.* For example, Pete's tit-for-tat strategy would be to confess the next time the game is played if Ann confesses in the current game, and not to confess the next time the game is played if Ann does not confess in the current game. A tit-for-tat strategy gives the other player incentive to follow the cooperative action—not confess—and thereby leads to a cooperative outcome. There are several other strategies that players can use to support a specific outcome in a repeated game.

Secret Defections

The incentive for one firm to defect from an agreement will depend on how likely it is that other firms will detect the defection. In the pumpkin example, it is impossible for Jack to increase his production without Jill's knowing it. This makes defection less likely. If one firm can secretly increase production or cut prices, enforcing the agreement will be more difficult. This is the problem with the world coffee cartel; it is

READING THE NEWS ABOUT... Oligopoly

The online video rental market, dominated by Netflix since its inception in 1999, has been challenged in recent years by the entry of Blockbuster and Wal-Mart. But as the article below implies, the exit of Wal-Mart might change this market situation. What do you think Mr. Squali, the analyst quoted in the article, means when he says that with the segment now a duopoly, "there is a chance for a price improvement and an end to the price bloodbath"?

Wal-Mart Ends Online Video Rentals and Promotes Netflix

By SAUL HANSELL (NYTimes)
May 20, 2005

Wal-Mart, which dominates so much of the retail world, is retreating from one of its more ambitious online ventures, a DVD rental service meant to compete in the market pioneered by Netflix. Instead, Wal-Mart said yesterday it had struck a deal to refer its online video rental customers to Netflix.

The withdrawal is another sign that Wal-Mart's power in brick-and-mortar retailing does not extend easily into the online world. Walmart.com, which sells more than a million individual items, has not proved to be a major threat to online leaders like Amazon.com. In contrast with its stores, Wal-Mart's online operation has not offered significantly lower prices than its rivals.

John Fleming, the company's executive vice president and chief marketing officer, who also oversees Walmart.com, said in a phone interview that the video rental service no longer fits into Wal-Mart's Internet strategy.

"The real big opportunity for us is in businesses that have the potential to integrate with our stores," he said.

In particular, he said that DVD sales on Walmart.com were growing rapidly, drawing strength from Wal-Mart's dominance of the overall DVD sales market. It accounts for about one-third of the DVD's sold in the United States.

Netflix and Wal-Mart declined to discuss the financial terms of the arrangement, but Netflix said it would not have a material effect on its financial results for the year.

Analysts said that Netflix would probably pay a bounty to Wal-Mart for each customer that converted to its service, much as it does with many other business partners. In addition, Netflix will promote Wal-Mart's DVD sales on its site, although analysts said this promotion would probably have little impact.

In the past, Netflix has had similar arrangements with Amazon.com and Best Buy, but those generated few sales.

Wal-Mart entered the online rental market in June 2003, and by March 31 of this year, it had attracted just under 300,000 customers, according to Majestic Research, a New York investment analysis firm.

easy to ship coffee around the world or cut prices without being detected. Crude oil shipments are more easily seen, but a member of OPEC could try to sell oil secretly to China. This might go on for a long time without detection. The impact of such secret defections is much like the situation in boxes 2 and 4 in Figure 11.6. Profits to the defector increase, and profits to the other producers decrease.

Netflix, by contrast, said it had three million subscribers on March 31. And Blockbuster, the dominant rental store chain, which entered the online rental market last August, had 820,000 subscribers, according to Majestic.

Blockbuster appears to have been able to gain traction against Netflix through a combination of low pricing, aggressive promotion and links with its video rental stores.

Blockbuster responded yesterday to the news of Wal-Mart's retreat by offering former Wal-Mart and Netflix customers a two-month free trial of its rental service.

Netflix appears to remain in the lead, analysts said, because its brand is so associated with online rentals and because it has more distribution centers, providing faster deliveries for customers.

All of the rental services allow users to rent a set number of DVD's at a time. As soon as customers mail back one DVD, they can receive another one from a list they enter on a Web site. Blockbuster, however, also gives its customers two free rentals a month from its stores, and Majestic Research has found that 60 percent of Blockbuster users download the free in-store rental coupon each month.

The competition among the rivals has led to a price war. Netflix initially offered the most popular version of its service—which allows three DVD's to be rented at one time—in 1999 for $19.99 a month, raising that price to $21.99 in early 2004. Wal-Mart's three-DVD plan had been $18.76 a month. But last November, when Blockbuster lowered its price to $17.49 a month, Netflix dropped its price to $17.99, and Wal-Mart to $17.36. In January, Blockbuster lowered its price again, to $14.99.

Netflix did not lower prices, but it started promoting a two-DVD plan for $14.99.

Both Blockbuster and Netflix plan to grow rapidly. Netflix said it hoped to end the year at approximately four million subscribers. Blockbuster said it hoped to have two million subscribers by the end of March 2006.

This war has been expensive for all sides. The average revenue per user a month at Netflix has dropped to $18.92 in the first quarter of 2005, from $22.51 in the third quarter of 2004, according to Youssef Squali, an analyst with Jefferies & Company, who does not own Netflix shares.

Blockbuster said it would invest $120 million in marketing and operations for the online rental service this year, in addition to the $50 million it spent last year. The service could break even next year, it said. Mr. Squali said that Wal-Mart's departure from the market could encourage both companies to raise prices. He pointed to the investor revolt at Blockbuster, which resulted in the election of the financier Carl C. Icahn to its board, as a sign the company might rethink its spending on the online rental market.

"Since the segment has become a duopoly," Mr. Squali said, "there is a chance for a price improvement and an end to the price bloodbath."

Shane Evangelist, general manager of Blockbuster online, said that for now his orders from the board remain the same.

"We are going to grow the business as fast as possible," he said.

For a long time, Japanese construction firms operated a now well-known collusion scheme called *dango*. All firms submitted high-priced bids to the government and took turns offering slightly lower bids. Ironically, and unfortunately for consumers, making the bids public made it harder for any firm to defect because firms in the agreement would know at once which firm had lowered its prices.

REVIEW

- Game theory provides a framework for studying strategic behavior in an oligopoly. Games, including the prisoner's dilemma, describe the strategies firms can use when they have the options of charging a monopoly price or a lower price.
- Game theory illustrates why firms in an oligopoly will be tempted to defect from any agreement.
- To the extent that a firm colludes, either explicitly or tacitly, it reduces economic efficiency by raising price above marginal cost.

Conclusion

In this chapter, we have explored two different types of models—monopolistic competition and oligopoly—that lie in the complex terrain between competition and monopoly. The models were motivated by the need to explain how real-world firms—Johnson Publications, Liz Claiborne, Nike, PepsiCo, the members of OPEC, and the members of the coffee cartel—operate in markets with differentiated products or with a small number of other firms.

In the models introduced in this chapter, firms have market power in that they can affect the price of the good in their market. Market power enables a firm to charge a price higher than marginal cost. It is a source of deadweight loss. Observations of the behavior of actual firms show a wide variation in market power among firms.

The ideas about monopolistic competition and oligopoly discussed in this chapter are used by economists in government and businesses. Economists working in the U.S. Department of Justice use them to determine whether the government should intervene in certain industries, as we will explore in Chapter 12. Consultants to business use them to help firms decide how to differentiate their products from those of other firms.

Having concluded our discussion of the four basic types of models of markets in this chapter, it is useful to remember the important distinction between *models* and the *facts* the models endeavor to explain or predict. None of the assumptions of these models—such as homogeneous products or free entry—hold exactly in reality. For example, when contrasted with the monopolistic competition model of this chapter, the model of competition, with its assumption of homogeneous goods, might seem not to apply to very many markets at all. Very few goods are exactly homogeneous. But when economists apply their models, they realize that these models are approximations of reality. How close an approximation comes to reality depends much on the application. The model of competition can be helpful in explaining the behavior of firms in industries that are approximately competitive, just as the model of monopoly can be helpful in explaining the behavior of firms in industries that are approximately monopolistic. Now we have a richer set of models that apply to situations far removed from competition or monopoly.

KEY POINTS

1. Firms that can differentiate their product and act strategically are in industries that fall between competition and monopoly.
2. Product differentiation—the effort by firms to create different products of value to consumers—is pervasive in a modern market economy. It helps explain intraindustry trade, advertising, and information services.

3. Monopolistic competition arises because of product differentiation. With monopolistic competition, firms have market power, but exit from and entry into the industry lead to a situation of zero profits in the long run.
4. With monopolistic competition, the firm sets the quantity produced so that price exceeds marginal cost. As a result, there is a deadweight loss, and average total cost is not minimized.
5. The deadweight loss and excess costs of monopolistic competition are part of the price paid for product variety.
6. Strategic behavior occurs in industries with a small number of firms because each firm has market power to affect the price, and each firm cannot ignore the response of other firms to its own actions.
7. Game theory suggests that noncooperative outcomes are likely, implying that collusive behavior will frequently break down, unless firms acquire a reputation for not defecting and secret defections can be prevented.

KEY TERMS

monopolistic competition
oligopoly
product differentiation
intraindustry trade
interindustry trade
excess costs
excess capacity
strategic behavior
game theory
prisoner's dilemma
payoff matrix
cooperative outcome
noncooperative outcome
Nash equilibrium
explicit collusion
cartel
tacit collusion
price leader

QUESTIONS FOR REVIEW

1. What is product differentiation?
2. What factors are relevant to the determination of optimal product differentiation?
3. Why is product differentiation an important reason for monopolistic competition?
4. What are two key differences between monopolistic competition and monopoly?
5. Why don't monopolistic competitors keep their average total cost at a minimum?
6. Why is the noncooperative outcome of a prisoner's dilemma game likely?
7. Why is duopoly like a prisoner's dilemma?
8. What is the difference between explicit and tacit collusion?
9. Why are secret defections a problem for cartels?
10. What are two alternative ways to assess the market power of firms in an industry?

PROBLEMS

1. Match the following characteristics with the appropriate models of firm behavior and explain the long-run efficiency (or inefficiency) of each.
   a. Many firms, differentiated product, free entry
   b. Patents, licenses, or barriers to entry; one firm
   c. Many firms, homogeneous product, free entry
   d. Few firms, strategic behavior
2. Consider Al's gasoline station, which sells Texaco at a busy intersection along with three other stations selling Shell, Conoco, and Chevron.
   a. Draw the marginal cost, average total cost, demand, and marginal revenue for Al's station, assuming that the profit-maximizing price is greater than average total cost. Show Al's profits.
   b. Explain what would happen in this situation to bring about a long-run equilibrium. Would more stations open, or would some leave?
3. Suppose the government places a sales tax on firms in a monopolistically competitive industry. Draw a diagram showing the short-run impact and the adjustment to the new long-run industry equilibrium. What happens to the equilibrium price and number of firms in the industry?
4. Compare the long-run equilibrium of a competitive firm with that of a monopolistically competitive firm with the same cost structure. Why is the long-run price different in these two models? Does the monopolistically competitive firm operate at a minimum cost? Draw a diagram and explain.
5. Suppose monopolistically competitive firms in the software industry make a technological improvement that shifts down average total cost but does not affect marginal cost. What will happen to the equilibrium number of firms, the quantity produced, and the long-run price of software?

6. Suppose there are 10 monopolistically competitive restaurants in your town with identical costs. Given the following information, calculate the short-run price and quantity produced by each of the firms.

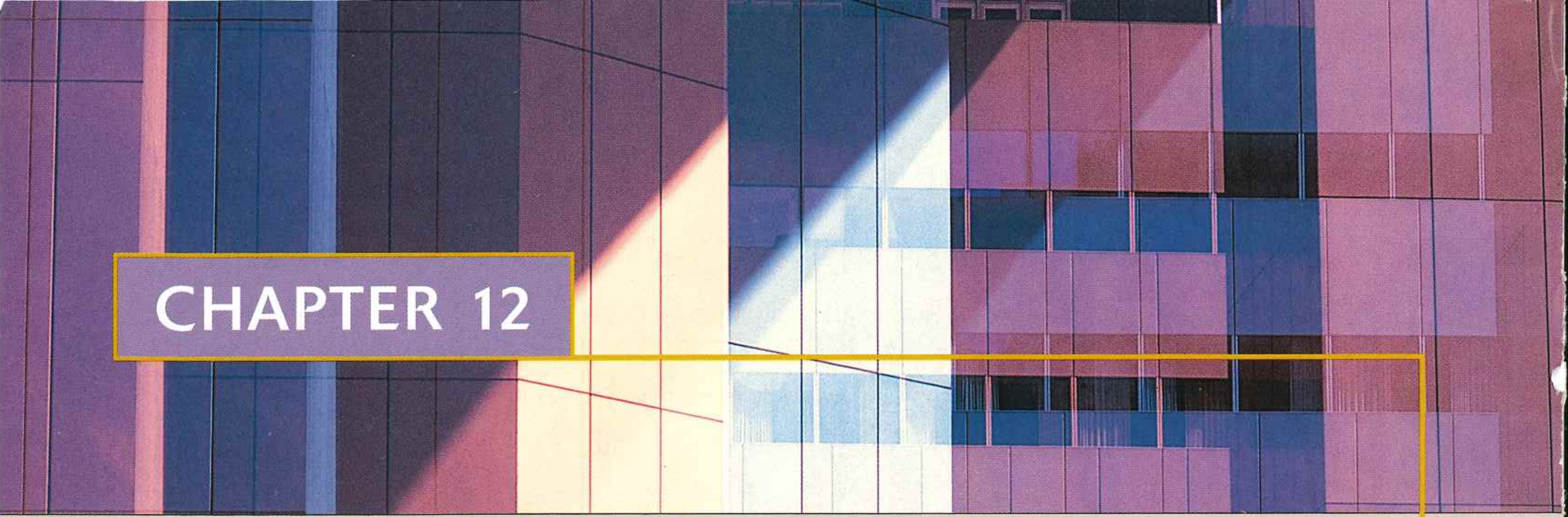
Each Firm's Demand		Each Firm's Costs	
Quantity	*Price*	*Average Total Cost*	*Marginal Cost*
1	10.00	13	—
2	8.00	9	5
3	6.00	8	6
4	4.00	9	12
5	2.00	10	14

a. Would the price rise or fall at the typical firm in the long run? Explain.
b. What would be the level of production if this industry were a competitive industry?
c. If there is free entry and exit in both monopolistic competition and competition, why is there a difference in the quantity the typical firm produces?

7. Which of the following conditions will tend to induce collusion among sellers in a market?
a. The transactions are publicly announced.
b. There are few sellers.
c. Some sellers have lower costs than other sellers.
d. The market is open for only one year.
e. The sellers cannot meet one another.

8. Two firms, Faster and Quicker, are the only two producers of sports cars on an island that has no contact with the outside world. The firms collude and agree to share the market equally. If neither firm cheats on the agreement, each firm makes $3 million in economic profits. If only one firm cheats, the cheater can increase its economic profit to $4.5 million, while the firm that abides by the agreement incurs an economic loss of $1 million. If both firms cheat, they earn zero economic profit. Neither firm has any way of policing the actions of the other.
a. What is the payoff matrix of the game that is played just once?
b. What is the equilibrium if the game is played only once? Explain.
c. What do you think will happen if the game can be played many times? Why?
d. What do you think will happen if a third firm comes into the market? Will it be harder or easier to achieve cooperation among the three firms? Why?

9. Store A and Store B are the only two flower shops in a small town. The demand for a dozen roses is $P = 25 - Q$. Neither Store A nor Store B has any fixed costs, whereas the marginal cost of Store A is constant at $3, and the marginal cost of Store B is constant at $5. Each seller can sell either 5 dozen or 10 dozen roses, and they meet only once in this market.
a. Create the payoff matrix. Show your calculations and explain verbally as necessary.
b. Find the Nash equilibrium or equilibria. Explain verbally.
c. If there are multiple equilibria, which equilibrium do you think is most likely to occur and why?



CHAPTER 12

Antitrust Policy and Regulation

When Microsoft came to dominate the personal computer software industry in the 1990s, the U.S. Department of Justice charged the company with using monopoly power to restrict competition, and Microsoft's founder, Bill Gates, was called before a federal judge to defend the company against the charges. The media compared Bill Gates's Microsoft monopoly with John D. Rockefeller's Standard Oil monopoly of the 1890s. When two office superstores, Staples and Office Depot, wanted to merge, the U.S. Federal Trade Commission objected and eventually succeeded in stopping the merger. When a new wireless communications device, the personal communications server, was developed, another government agency, the Federal Communications Commission, determined in advance the very structure of the new market—how many firms there would be in each region and whether or not existing cellular phone companies could compete.

These events represent just a few of the thousands of ways in which the government intervenes in the operations of firms. The intent of the government in many of these cases is to promote competition, which we know is an essential ingredient of market efficiency. Recall that the models of monopoly and monopolistic competition in Chapters 10 and 11 show that when firms have market power, they raise prices above marginal cost, reduce the quantity produced, and create a deadweight loss to society. In such cases, the government may be able to intervene to reduce the deadweight loss and increase economic efficiency.

This chapter uses the models developed in Chapters 10 and 11 to explain the different ways the government can promote competition and regulate firms with market power. We consider two broad types of policy: (1) antitrust policy, which is concerned with preventing anticompetitive practices like price fixing and with limiting firms' market power by preventing mergers or breaking up existing firms, and (2) regulatory policy, in which the government requires firms that have a natural monopoly to set prices at prescribed levels.

Antitrust Policy

antitrust policy: government actions designed to promote competition among firms in the economy; also called competition policy or antimonopoly policy.

Antitrust policy refers to the actions the government takes to promote competition among firms in the economy. Antitrust policy includes challenging and breaking up existing firms with significant market power, preventing mergers that would increase monopoly power significantly, prohibiting price fixing, and limiting anticompetitive arrangements between firms and their suppliers.

Attacking Existing Monopoly Power

Antitrust policy began in the United States just over 100 years ago in response to a massive wave of mergers and consolidations. Similar merger movements occurred in Europe at about the same time. These mergers were made possible by rapid innovations in transportation, communication, and management techniques. Railroads and telegraph lines expanded across the country, allowing large firms to place manufacturing facilities and sales offices in many different population centers. It was during this period that the Standard Oil Company grew rapidly, acquiring about 100 firms and gaining about 90 percent of U.S. oil refinery capacity. Similarly, the United States Steel Corporation was formed in 1901 by merging many smaller steel companies. It captured about 65 percent of the steel ingot market. These large firms were called *trusts*.

Sherman Antitrust Act: a law passed in 1890 in the United States to reduce anticompetitive behavior; Section 1 makes price fixing illegal, and Section 2 makes attempts to monopolize illegal.

The **Sherman Antitrust Act** of 1890 was passed in an effort to prevent these large companies from using their monopoly power. Section 2 of the act focused on the large existing firms. It stated, "Every person who shall monopolize, or attempt to monopolize . . . any part of the trade or commerce among the several states, or with foreign nations, shall be deemed guilty of a felony."

■ **A Brief History: From Standard Oil to Microsoft.** It was on the basis of the Sherman Antitrust Act that Theodore Roosevelt's administration took action to break apart Standard Oil. After 10 years of litigation, the Supreme Court ruled in 1911 that Standard Oil monopolized the oil-refining industry illegally. To remedy the problem, the courts ordered that Standard Oil be broken into a number of separate entities. Standard Oil of New York became Mobil; Standard Oil of California became Chevron; Standard Oil of Indiana became Amoco; Standard Oil of New Jersey became Exxon. Competition among these companies was slow to develop, since their shares were still controlled by Rockefeller. But as the shares were distributed to heirs and then sold, the companies began to compete against each other. Now the oil-refining companies have much less monopoly power. In fact, with the greater degree of competition, the Clinton administration began to allow some of these firms to merge, although not into one single oil-refining firm. For example, on November 30, 1999, Exxon and Mobil merged to form a new firm called Exxon Mobil.

Soon after its success in splitting apart Standard Oil, the U.S. government took successful action under the Sherman Act against the tobacco trust, splitting up the American Tobacco Company into sixteen different companies. It also broke up several monopolies in railroads, food processing, and chemicals. However, the government was not successful in using the Sherman Act against United States Steel. As part of the Standard Oil decision, the Supreme Court developed a **rule of reason** that required not only that a firm have monopoly power but also that it intend to use that power against other firms in a way that would restrict competition. Monopoly *per se*, in and of itself, was not enough, according to the Supreme Court in 1911. Since most competitors and customers of United States Steel said that the company's actions did not restrain competition, the Supreme Court, applying its rule of reason, decided in 1920 that United States Steel was not guilty under the Sherman Act.

rule of reason: an evolving standard by which antitrust cases are decided, requiring not only the existence of monopoly power but also the intent to restrict trade.

Twenty-five years later, a 1945 Supreme Court decision that found Alcoa Aluminum guilty of monopolization refined the rule of reason to make it easier to prove guilt. Although a monopoly per se was still not enough, the intent to willingly acquire and maintain a monopoly—easier to prove than an intent to restrict competition—was enough to establish guilt.

In 1969 the U.S. government brought antitrust action against IBM because of its dominance in the mainframe computer market. After a number of years of litigation, the government dropped the case. One reason was rapid change in the computer market. Mainframes were facing competition from smaller computers. Firms such as Digital Equipment and Apple Computer were competing with IBM by 1982, when the government withdrew its case. Looking at the competition picture more broadly and recognizing that it had already spent millions, the government decided that antitrust action was no longer warranted.

The U.S. government took action against AT&T in the 1970s. It argued that AT&T, as the only significant supplier of telephone service in the nation, was restraining trade. As a result of that antitrust action, AT&T was broken apart and had to compete with MCI and Sprint in providing long-distance telephone service nationwide. This increase in competition lowered the cost of long-distance calls.

The most recent big case was brought against Microsoft. After several antitrust-related investigations, negotiations, and lawsuits in the 1990s, a federal judge found that Microsoft had monopoly power and used it to harm its competitors and consumers, and ordered the firm's breakup in June 2000. However, that order was reversed in 2001, and the federal government reached an agreement whereby Microsoft would provide over a billion dollars in computer software and services and cash to public schools. Several state governments opposed the agreement, seeking stronger penalties on Microsoft.

■ **Predatory Pricing.** Attempts by firms to monopolize by predatory pricing have also been challenged by the government and by other firms, though breakup is not usually the intended remedy. **Predatory pricing** refers to the attempt by a firm to charge a price below its shutdown point in order to drive its competitors out of business, after which it then forms a monopoly.

predatory pricing: action on the part of one firm to set a price below its shutdown point in order to drive its competitors out of business.

A 1986 Supreme Court decision, *Matsushita v. Zenith*, has made predatory pricing harder to prove. Matsushita and several other Japanese companies were accused by Zenith of predatory pricing of televisions in the U.S. market. After five years of litigation and appeals, the Court decided that there was not sufficient evidence for predatory pricing. The Court argued that the Japanese firms' share of the U.S. market was too small compared to Zenith's to make monopolization plausible. Moreover, the low price of the Japanese televisions seemed to be based on low production costs. Thus, the Court's majority opinion stated that this predatory pricing case appeared to make "no economic sense."

"This town isn't big enough for both of us— let's merge."

Predatory pricing is difficult to distinguish from vigorous competition, which is essential to a well-functioning market economy. For example, Wal-Mart has been accused of predatory pricing by smaller retailers, who find it is hard to compete with Wal-Mart's low prices. Yet, in many of these cases, it is likely that Wal-Mart is more efficient. Its lower prices are due to lower costs. In 1993, Northwest Airlines sued American Airlines for predatory pricing in Texas but lost. The jury decided that although American Airlines was charging prices below its shutdown point, it was not attempting to monopolize the market.

Merger Policy

There were thirty-three breakups of firms ordered by the courts from 1890 to 1981, including those of Standard Oil and AT&T. However, there have been no breakups since the AT&T breakup in 1981. There has been a decline in the frequency of government-forced breakups in recent years, which may be due to greater international competition or to the effectiveness of merger policy, which we now consider. For firms to occupy a huge share of the market, they must either grow internally or merge with other firms. A merger policy that prevents mergers that create firms with huge market power reduces the need to break up firms.

The Legislation, the Antitrust Division, and the FTC. The Sherman Antitrust Act dealt with monopolies that were already in existence. The **Clayton Antitrust Act** of 1914 aimed to prevent the creation of monopolies and now provides the legal basis for preventing mergers that would significantly reduce competition. The **Federal Trade Commission (FTC)** was set up in 1914 to help enforce these acts along with the Justice Department.

To this day, the **Antitrust Division of the Justice Department** and the FTC have dual responsibility for competition policy in the United States. The Justice Department has more investigative power and can bring criminal charges, but for the most part, there is a dual responsibility.

Clayton Antitrust Act: a law passed in 1914 in the United States aimed at preventing monopolies from forming through mergers.

Federal Trade Commission (FTC): the government agency established to help enforce antitrust legislation in the United States; it shares this responsibility with the Antitrust Division of the Justice Department.

Antitrust Division of the Justice Department: the division of the Justice Department in the United States that enforces antitrust legislation, along with the Federal Trade Commission.

Herfindahl-Hirschman index (HHI): an index ranging in value from 0 to 10,000 indicating the concentration in an industry; it is calculated by summing the squares of the market shares of all the firms in the industry.

Economic Analysis. How does the government decide whether a merger by firms reduces competition in the market? The economists and lawyers in the Justice Department and the FTC provide much of the analysis. They focus on the market power of the firm. The more concentrated the firms in an industry, the more likely it is that the firms have significant market power. Concentration is usually measured by the Herfindahl-Hirschman index.

The "Herf." The **Herfindahl-Hirschman index (HHI)** is used so frequently to analyze mergers that it has a nickname: the "Herf." The HHI is defined as the sum of the squares of the market shares of all the firms in the industry. The more concentrated the industry, the larger the shares and, therefore, the larger the HHI. For example, if there is one firm, the HHI is $(100)^2 = 10{,}000$, the maximum value. If there are two firms, each with a 50 percent share, the HHI is $(50)^2 + (50)^2 = 5{,}000$. If there are 10 firms with equal shares, the HHI is 1,000. Values of the HHI for several hypothetical examples of firm shares in particular industries are listed in Table 12.1.

Observe that the HHI tends to be lower when there are more firms in the industry and when the shares of each firm are more equal. Even when the number of firms in the industry is very large, the HHI can be large if one or two firms have a large

Table 12.1
Examples of the HHI in Different Industries

Industry Example	Number of Firms	Shares (percent)	HHI
A	3	42,42,16	3,784
B	4	42,42,8,8	3,656
C	5	42,42,8,4,4	3,624

share. For example, an industry with 20 firms in which one firm has 81 percent of the market and the others each have 1 percent has a very large HHI of 6,580, even greater than that of a two-firm industry with equal shares.

According to the *merger guidelines* put forth by the Justice Department and the FTC, mergers in industries with a postmerger HHI *above 1,800* will likely be challenged if the HHI rises by 50 points or more. When the HHI is *below 1,000*, a challenge is unlikely. *Between 1,000 and 1,800*, a challenge will likely occur if the HHI rises by 100 points or more.

Some examples are found in Table 12.1. Suppose that the two smallest firms in industry C in Table 12.1 merge and the industry thereby takes the form of industry B. Then the HHI rises by 32, from 3,624 to 3,656. Hence it is unlikely that the government would challenge this merger. In contrast, suppose that the two smallest firms in industry B merge. Then the HHI increases by 128 and the government would be likely to challenge the merger.

The HHI is used because it indicates how likely it is that firms in the industry after the merger will have enough market power to raise prices well above marginal cost, reduce the quantity produced, and cause economic inefficiency. For example, when the FTC blocked the merger of Office Depot and Staples in 1997, it stated that the "post-merger HHIs average over 3000" and that "increases in HHIs are on average over 800 points."[1]

The FTC or Justice Department looks at other things in addition to concentration measures. Ease of entry of new firms into the industry is an important factor, as is the potential contestability of the market by other firms. Recall the idea of *contestable markets* discussed in Chapter 10: Even if firms are highly concentrated in an industry, potential entry by other firms provides competitive pressure on the industry. Thus, an industry with a high degree of concentration may, in fact, be acting competitively because of the threat of new firms coming into the business.

contestable market: a market in which the threat of competition is enough to encourage firms to act like competitors. (Ch. 10)

Market Definition. When measuring concentration in a market, the market definition is very important. **Market definition** is a description of the types of goods and services included in the market and the geographic area of the market. Table 12.2 shows the range of possibilities for market definition when considering the merger of soft drink producers. Should the market definition be narrow (carbonated soft drinks) or broad (all nonalcoholic beverages)? The market definition makes a big difference for concentration measures. In 1986 the FTC blocked a merger between Coca-Cola and Dr. Pepper, which would have increased the HHI by 341 in the carbonated soft drink market. In contrast, the HHI would have increased by only 74 if bottled water, powdered soft drinks, tea, juices, and coffee were also included in the market, along with carbonated soft drinks.

market definition: demarcation of a geographic region and a category of goods or services in which firms compete.

[1] Public Brief to D.C. District Court on *FTC* v. *Staples and Office Depot*, April 7, 1997.

Table 12.2
Different Market Definitions in the Beverage Industry

						Milk
					Tea	Tea
				Coffee	Coffee	Coffee
			Juice drinks	Juice drinks	Juice drinks	Juice drinks
		Bottled water	Bottled water	Bottled water	Bottled water	Bottled water
	Powdered soft drinks	Powdered soft drinks	Powdered soft drinks	Powdered soft drinks	Powdered soft drinks	Powdered soft drinks
Carbonated soft drinks	Carbonated soft drinks	Carbonated soft drinks	Carbonated soft drinks	Carbonated soft drinks	Carbonated soft drinks	Carbonated soft drinks
Narrow Market Definition			**Medium Market Definition**			**Broad Market Definition**

Defining the geographic area of a market is also a key aspect of defining the market for a good or service. In an integrated world economy, a significant amount of competition comes from firms in other countries. For example, in the automobile industry in the United States, there have been only three major producers. This is a highly concentrated industry. However, intense competition coming from Japanese, Korean, German, and other automobile companies increases the amount of competition. The rationale for challenging a merger is mitigated substantially by international competition.

■ **Horizontal versus Vertical Mergers.** Merger policy also distinguishes between **horizontal mergers,** in which two firms selling the same good or the same type of good merge, and **vertical mergers,** in which a firm merges with its supplier, as, for example, when a clothing manufacturer merges with a retail clothing store chain. The merger guidelines refer to horizontal mergers. Virtually all economists agree that horizontal mergers have the potential to increase market power, all else the same.

horizontal merger: a combining of two firms that sell the same good or the same type of good.

vertical merger: a combining of two firms, one of which supplies goods to the other.

There is considerable disagreement among economists about the effects of vertical mergers, however. A vertical merger will seldom reduce competition if there are firms competing at each level of production. However, some feel that a vertical merger may aid in reducing competition at the retail store level.

Price Fixing

In addition to breaking up firms and preventing firms with a great amount of market power from merging, antitrust policy looks for specific forms of conspiracy to restrict competition among firms. For example, when two or more firms conspire to fix prices, they engage in an illegal anticompetitive practice. **Price fixing** is a serious, frequently criminal, offense. Section 1 of the Sherman Antitrust Act makes price fixing illegal *per se.*

price fixing: the situation in which firms conspire to set prices for goods sold in the same market.

READING THE NEWS ABOUT... Proposed Mergers

Staples and Office Depot announced plans to merge in 1996. At the time of the announcement, the two firms teamed up to place a full-page advertisement in major newspapers around the country. In huge print the ad stated:

> **Something Special Will Happen When**
> **the Two Low Price Leaders Combine . . .**
> LOWER PRICES
> **On Office Products Every Day!**

The ad then went on to explain how the proposed merger would lower prices by reducing costs through economies of scale. Through this ad, the companies were making their own news, trying to influence public opinion and thereby get approval for the merger from the Federal Trade Commission (FTC).

But the FTC did not believe the ads. The economists at the FTC immediately found that the HHI would increase by large amounts if one defined the market as "office superstores" and excluded stores such as Wal-Mart that also sold office supplies. The FTC also found that prices for paper, ballpoint pens, envelopes, and so on, were higher, not lower, in areas with only one store than in areas where Office Depot and Staples competed. So, the FTC asked a federal judge to prevent the merger. The FTC claimed that "the proposed merger would violate Section 7 of the Clayton Act."

Economists served as expert witnesses on both sides of the issue, arguing about whether to define the market narrowly or broadly and about whether prices would be lower or higher with the merger. The judge was not convinced by the two companies, by their expert witnesses, or by the ads. The judge ruled to stop the merger. On the question of market definition, the judge argued in his opinion that "No one entering a Wal-Mart would mistake it for an office superstore. No one entering Staples or Office Depot would mistakenly think he or she was in Best Buy or CompUSA. You certainly know an office superstore when you see one." If you were the judge, would you have included Wal-Mart and those other firms in the definition of the office supply market? Would you be skeptical about such ads if you saw them in the newspaper now?

Laws against price fixing are enforced by bringing lawsuits against the alleged price fixers. Suits are brought both directly by the Justice Department and by individual firms that are harmed by price fixing. The number of private suits greatly exceeds the number of government suits. Individual firms can collect **treble damages** (a provision included in the Clayton Act)—three times the actual damages. The treble damage penalty aims to deter price fixing.

treble damages: penalties awarded to the injured party equal to three times the value of the injury.

One of the most famous price-fixing cases in U.S. history occurred in the 1950s and involved Westinghouse and General Electric. Through an elaborate system of secret codes and secret meeting places, the executives of these two firms agreed together to set the price of electrical generators and other equipment they were selling in the same market. Through this agreement, they set the price well above competitive levels, but they were discovered and found guilty of price fixing. Treble damages amounting to about $500 million were awarded, and criminal sentences were handed down; some executives went to prison.

A more recent price-fixing case involved the production of food additives. The large agricultural firm Archer-Daniels-Midland (ADM) was sued by the Justice Department for fixing prices with other international producers. In 1996, as part of the settlement in this case, ADM paid over $100 million in fines.

Table 12.3
Price-Cost Margins in Several Industries

Industry	Price-Cost Margin
Food processing	.50
Coffee roasting	.04
Rubber	.05
Textiles	.07
Electrical machinery	.20
Tobacco	.65
Retail gasoline	.10
Standard automobiles	.10
Luxury automobiles	.34

Source: T. F. Bresnahan, "Empirical Studies of Industries with Market Power," *Handbook of Industrial Organization*, Vol. II, ed. R. Schmalensee and R. D. Willig (Amsterdam: Elsevier Science Publishers, 1989).

■ **Price-Cost Margins.** A way of measuring market power is the *price-cost margin*. The greater the price (*P*) is above the marginal cost (*MC*), the more market power firms have. Table 12.3 gives some estimates of the price-cost margin $[(P - MC)/P]$ for firms in several different industries. The higher the price-cost margin, the more market power firms in the industry have. Observe in Table 12.3 that the price-cost margin is very small for coffee roasting, rubber, textiles, retail gasoline, and standard automobiles. The firms in these markets apparently have little market power. In contrast, the price-cost margin is very high for food processing and tobacco.

An interesting example is Anheuser-Busch, the producer of Budweiser beer. Before the introduction of Lite Beer by Miller, Anheuser-Busch had considerable market power; the price-cost margin was .3. After Lite Beer was introduced, the firm lost market power. The price-cost margin dropped to .03. Evidently Lite Beer made Miller a more visible player in the beer market and thus increased competition in the market in the sense that Anheuser-Busch's market power declined.

Vertical Restraints

The price-fixing arrangements just described are an effort to restrict trade in one horizontal market, such as the electrical machinery market or the market for food additives. Such restraints of trade clearly raise prices, reduce the quantity produced, and cause deadweight loss. But there are also efforts by firms to restrain trade vertically. For example, **exclusive territories** occur when a manufacturer of a product gives certain retailers or wholesalers exclusive rights to sell the product in a given area. This practice is common in soft drink and beer distribution. **Exclusive dealing** is the practice by which a manufacturer does not allow a retailer to sell goods made by a competitor. **Resale price maintenance** is the practice of a manufacturer's setting a list price for a good and then forbidding the retailer to offer a discount.

exclusive territories: the regions over which a manufacturer limits the distribution or selling of its products to one retailer or wholesaler.

exclusive dealing: a condition of a contract by which a manufacturer does not allow a retailer to sell goods made by a competing manufacturer.

resale price maintenance: the situation in which a producer sets a list price and does not allow the retailer to offer a discount to consumers.

Do vertical restraints reduce economic efficiency? There is considerable agreement among economists that manufacturers cannot increase their own market power by restraints on the firms to which they supply goods. A manufacturer's requiring that a retailer take a certain action does not give the manufacturer a greater ability to raise prices over competitors without losing sales. In addition, in some circumstances such restraints may actually increase economic efficiency.

Consider resale price maintenance, for example. Suppose that low-price discount stores compete with high-price retail stores that provide services to customers. If a discount store could offer the same product with little or no service, then people could go to the higher-price store, look the product over, get some useful advice from knowledgeable salespeople, and then buy at the discount store. Soon such services would disappear. Resale price maintenance can thus be viewed as a means of preserving such service by preventing the discount store from charging a lower price.

If a producer and retailer are vertically integrated into one firm, then clearly they coordinate the price decisions. For example, the Gap sells its own products in its retail outlets, and it obviously sets the retail price. Outlawing resale price maintenance would mean that firms that were not vertically integrated could not do the same thing as the Gap does. Why should Levi Strauss not be permitted to set the price of Levis sold at retail stores that compete with the Gap?

However, some argue that resale price maintenance is a way to reduce competition at the retail level. They see retailers having competitive pressure to keep prices low as more important than the possible loss of some retail customer services.

In sum, there is more controversy among economists about the effect of vertical restraints than about horizontal restraints.

REVIEW

- Breaking up monopolies, preventing mergers that would create too much market power, and enforcing laws against price fixing are the main government actions that constitute antitrust policy.
- Section 1 of the Sherman Antitrust Act outlaws price fixing.
- Section 2 of the Sherman Antitrust Act allows the government to break up firms with monopoly power.
- The Clayton Antitrust Act provides the legal basis for merger policy.
- All these policies aim to increase competition and thus improve the efficiency of a market economy.
- There is more controversy about the effects of vertical mergers and vertical restraints than about horizontal mergers and horizontal restraints.

Regulating Natural Monopolies

The goal of antitrust policy is to increase competition and improve the efficiency of markets. Under some circumstances, however, antitrust policy against a monopoly is not necessarily in the interest of efficiency. In the provision of certain goods, such as water, it is inefficient for more than one company to deliver the product to households. To provide its services, a water company must dig up the streets, lay the water pipes, and maintain them. It would be inefficient to have two companies supply water because that would require two sets of pipes and would be a duplication of resources. Another example is electricity. It makes no sense to have two electric utility firms supply the same neighborhood with two sets of wires. A single supplier of electricity is more efficient.

Economies of Scale and Natural Monopolies

natural monopoly: a single firm in an industry in which average total cost is declining over the entire range of production and the minimum efficient scale is larger than the size of the market. (Ch. 10)

Water and electricity are examples of *natural monopolies,* industries in which one firm can supply the entire market at a lower cost than two firms can. Recall from the discussion in Chapter 10 that the key characteristic of a natural monopoly is a declining average total cost curve. Average total cost declines as more is produced because fixed costs are very large compared to variable costs. Once the main line is laid for the water supply, it is relatively easy to hook up another house. Similarly, with electricity, once the main lines are installed, it is relatively easy to run wires into a house. A large initial outlay is necessary to lay the main water pipes or main electrical lines, but thereafter the cost is relatively low. The more houses that are hooked up, the less the average total cost is. Recall that when the long-run average total cost curve declines, there are *economies of scale.*

economies of scale: a situation in which long-run average total cost declines as the output of a firm increases. (Ch. 8)

Figure 12.1 shows graphically why one firm can always produce more cheaply than two or more firms when the average total cost curve is downward-sloping. The figure shows quantity produced on the horizontal axis and dollars on the vertical axis; a downward-sloping average total cost curve is plotted. If two firms divide up the market (for example, if two water companies supply water to the neighborhood), then the average total cost is higher than if one firm produces for the entire market. It

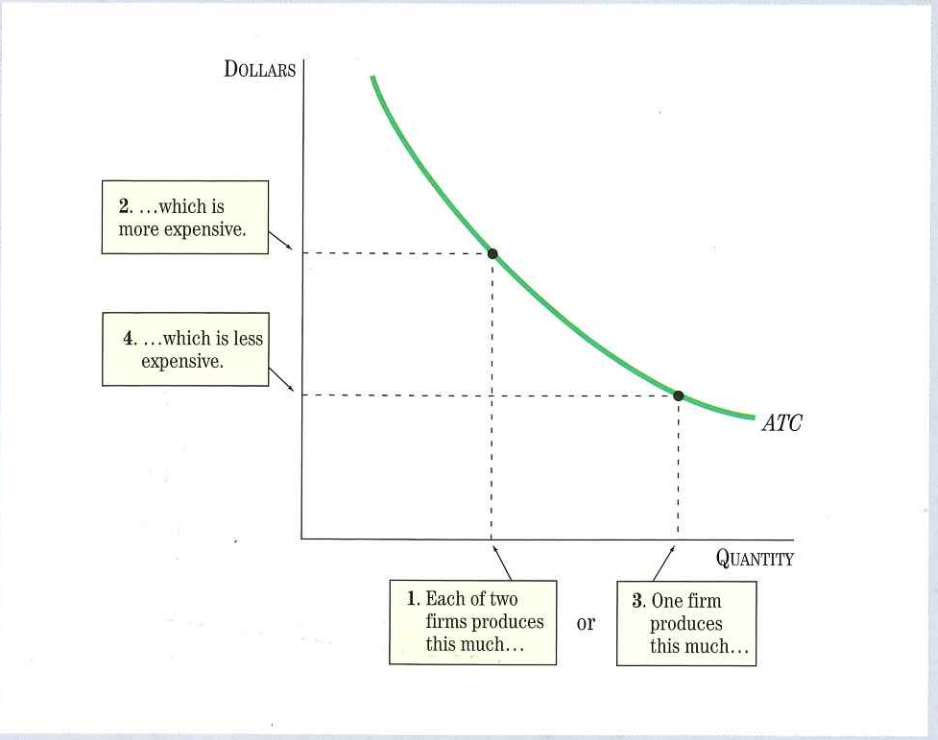

DOLLARS
2. ...which is more expensive.
4. ...which is less expensive.
ATC
QUANTITY
1. Each of two firms produces this much...
or
3. One firm produces this much...

Figure 12.1
Natural Monopoly: Declining Average Total Cost
If two firms supply the market, dividing total production between them, costs are higher than if one firm supplies the market. The costs would be even greater if more than two firms split up the market.

is more costly for two or more firms to produce a given quantity in the case of a declining average total cost curve than for one firm.

Alternative Methods of Regulation

What is the best government policy toward a natural monopoly? Having one firm in an industry lowers the cost of production, but there will be inefficiencies associated with a monopoly: Price will be higher than marginal cost, and there will be a deadweight loss. To get both the advantages of one firm producing *and* a lower price, the government can regulate the firm.

The monopoly price and quantity of a natural monopoly with declining average total cost are illustrated in Figure 12.2. The monopoly quantity occurs where marginal revenue equals marginal cost, the profit-maximizing point for the monopolist. The monopoly price is above marginal cost. If the firm's price was regulated, then the government could require the firm to set a lower price, thereby raising output and eliminating some of the deadweight loss associated with the monopoly. There are three ways for the government to regulate the price: marginal cost pricing, average total cost pricing, and incentive regulation.

Marginal Cost Pricing. We know that there is no deadweight loss with competition because firms choose a quantity of output such that marginal cost is equal to price. Hence, one possibility is for the government to require the monopoly to set its price equal to marginal cost. This method is called **marginal cost pricing.** However, with declining average total cost, the marginal cost is lower than average total cost. This is shown in Figure 12.2 for the case where marginal cost is constant. Thus, if price were equal to marginal cost, *the price would be less than average total cost,* and

marginal cost pricing: a regulatory method that stipulates that the firm charge a price that equals marginal cost.

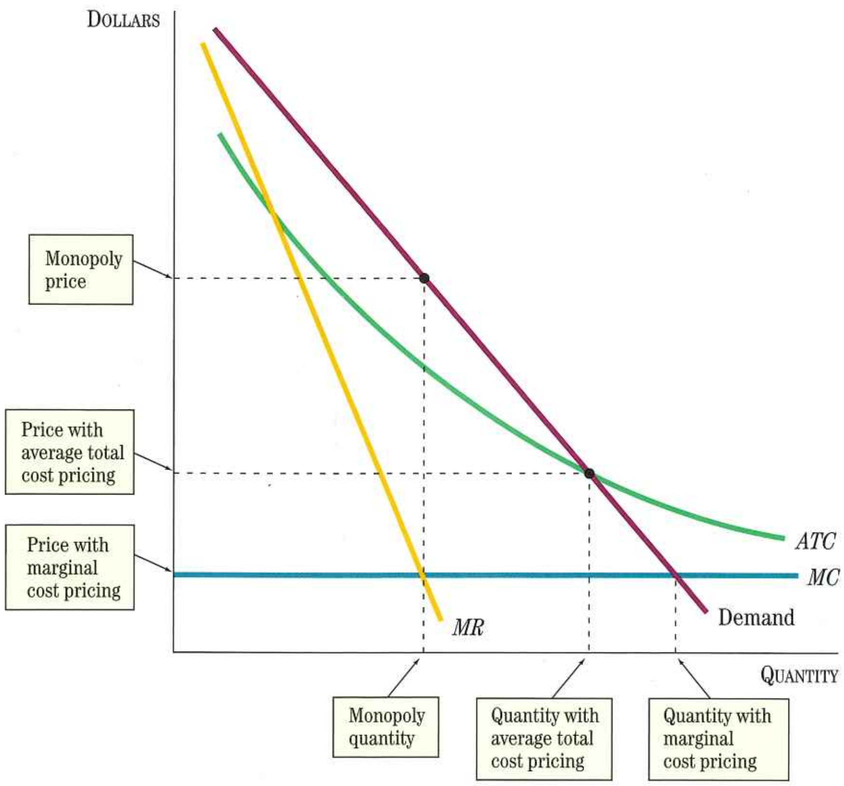


Figure 12.2
Monopoly Price versus Alternative Regulatory Schemes
Two alternatives, marginal cost pricing and average total cost pricing, are compared with the monopoly price. Marginal cost pricing gives the greatest quantity supplied, but since price is less than average total cost, the firm earns negative profits. Average total cost pricing results in a larger quantity supplied, and the firm earns zero economic profits.

the monopoly's profits would be negative (a loss). There would be no incentive for any firm to come into the market.

For example, if the regulators of an electrical utility use a pricing rule with price equal to marginal cost, there will be no incentive for the electrical utility to build a plant or produce electricity. Although the idea of mimicking a competitive firm by setting price equal to marginal cost might sound reasonable, it fails to work in practice.

average total cost pricing: a regulatory method that stipulates that the firm charge a price that equals average total cost.

■ **Average Total Cost Pricing.** Another method of regulation would have the firm set the price equal to average total cost. This is called **average total cost pricing** or, sometimes, cost-of-service pricing. It is also illustrated in Figure 12.2. When price is equal to average total cost, we know that economic profits will be equal to zero. With the economic profits equal to zero, there will be enough to pay the managers and the investors in the firm their opportunity costs. Although price is still above

ECONOMICS IN ACTION

California Electricity Crisis

Electricity has long been considered a natural monopoly, subject to federal regulation. In the United States, the Federal Power Act of 1935 led to the creation of a group of geographically contained "vertical" monopolies over the generation, distribution, and sale of power in each company's region. But in the 1990s, deregulation legislation allowed for market competition among power generators (though not among electricity distributors). In 2000 and 2001, California energy suppliers falsified price data and trading records to manipulate the Californian market and illegally increase the price of electricity and natural gas. The energy shortages forced California to ration electricity and use selective power outages. Some people believe that deregulation may have been partly responsible for California's electricity crisis, whereas others believe that the way partial deregulation was implemented—that is, the new set of rules and incentives in the market—were partially responsible for the crisis.

View down San Francisco's Market Street at the end of day 14 of a stage-three State of California power alert, in late January 2001.

marginal cost, it is less than the monopoly price; the deadweight loss will be smaller and more electricity will be produced compared with the monopoly.

But there are some serious problems with average total cost pricing. Suppose the firm knows that whatever its average total cost is, it will be allowed to charge a price equal to average total cost. In that situation, there is no incentive to reduce costs. Sloppy work or less innovative management could increase costs. With the regulatory scheme in which the price equals average total cost, the price would rise to cover any increase in cost. Inefficiencies could occur with no penalty whatsoever. This approach provides neither an incentive to reduce costs nor a penalty for increasing costs at the regulated firm.

■ **Incentive Regulation.** The third regulation method endeavors to deal with the problem that average total cost pricing provides too little incentive to keep costs low. The method is called **incentive regulation.** It is a relatively new idea, but it is quickly spreading, and most predict that it is the way of the future. The method projects a regulated price out over a number of years. That price can be based on an estimate of average total cost. The regulated firm is told that the projected price will not be revised upward or downward for a number of years. If the regulated firm achieves an average total cost lower than the price, it will be able to keep the profits, or perhaps pass on some of the profits to a worker who came up with the idea for the innovation. Similarly, if sloppy management causes average total cost to rise, then profits will fall because the regulatory agency will not revise the price.

incentive regulation: a regulatory method that sets prices for several years ahead and then allows the firm to keep any additional profits or suffer any losses over that period of time.

Thus, under incentive regulation, the regulated price is only imperfectly related to average total cost. The firm has a profit incentive to reduce costs. If a firm does poorly, it pays the penalty in terms of lower profits or losses.

Under incentive regulation, the incentives can be adjusted. For example, the California Public Utility Commission (the regulators of utility firms in California) has

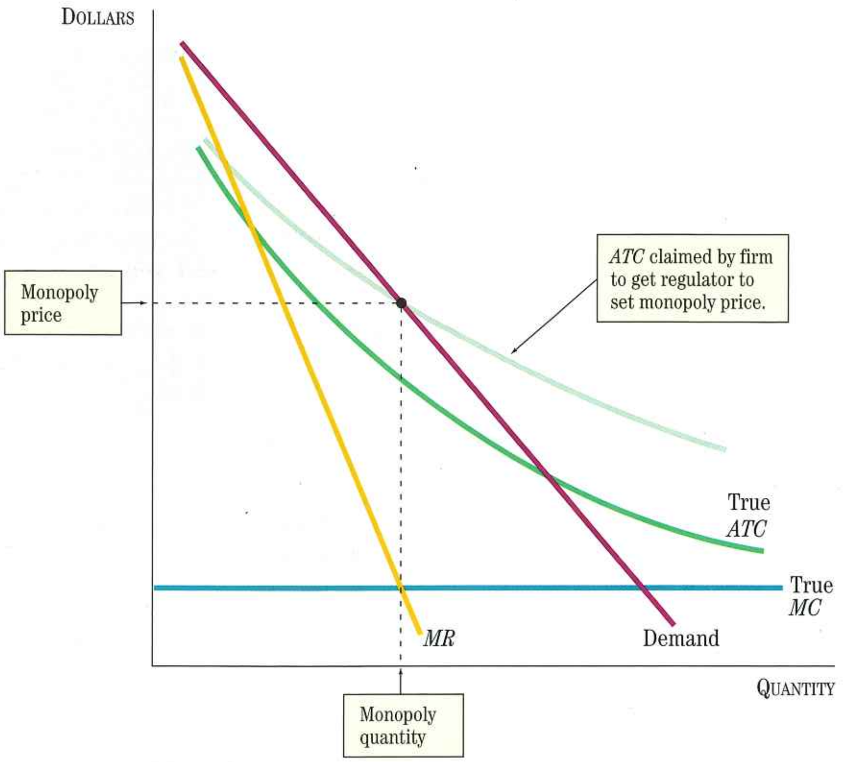


Figure 12.3
Asymmetric Information and Regulation
If a regulator uses average total cost pricing but does not have complete information about costs at the firm, the firm could give misleading information about its costs in order to get a higher price from the regulator. In an extreme case, shown in this figure, the firm could say its costs were so high that it could get the monopoly price.

incentive schemes by which electrical utility firms and their customers share equally in the benefits of reduced costs and in the penalties from increased costs. This reduces the incentive to the firm in comparison to the case where the benefits and penalties are not shared.

Incentive regulation is sometimes made difficult by *asymmetric information*; that is, when one of the parties has access to more or better information. In this case, the regulated firm has more information than the regulator about its equipment, technology, and workers. Thus, the firm can mislead the regulator and say that its average total cost is higher than it actually is in order to get a higher price, as shown in Figure 12.3.

REVIEW

- In the case of a natural monopoly, one firm can produce at a lower average total cost than two or more firms, but a monopoly causes deadweight loss.
- If government regulates the monopoly through marginal cost pricing, the firm will run losses.
- Average total cost pricing leads to increased costs because it doesn't provide incentives to keep costs down.
- Incentive regulation is becoming the preferred method of regulation.

To Regulate or Not to Regulate

Our analysis thus far suggests that the government should regulate firms' prices in situations where natural monopolies exist. In practice, this requires deciding when a natural monopoly exists, which is frequently difficult.

There are many examples in American history of the government's regulating a firm's prices even when it is far-fetched to think about the firm as a natural monopoly. For example, for a long period of time, trucking was regulated by the federal government. Trucking regulation grew out of railroad regulation, which was originally justified when railroads were the only rapid form of transportation and thus were natural monopolies. Under trucking regulation, the federal government put a floor on the price that trucking companies could charge when shipping goods interstate. Federal regulation of trucking was disbanded in the early 1980s. Studies have shown that trucking rates fell as a result.

Borderline Cases

Clearly, trucking is not a natural monopoly. The trucking industry is at the opposite end of the spectrum from water or electrical utility companies, which are almost always regulated.

But there are many borderline cases that are more controversial. Many of these arise in high-technology industries such as telecommunications and computing. An important example is cable television. In 1992, there was considerable debate about whether the federal government should regulate cable television. On first thought, it may appear that cable television is no different from electricity or water. Once a cable television company lays the cable down in a neighborhood, there is a fairly small cost to connect each individual house to it. On the other hand, there are alternatives to cable television for many homes. For example, over-the-air television channels do provide some competition to cable television. If one lives in an area where there are few over-the-air channels, there is little competition. However, if there are six, seven, or eight over-the-air channels, then there is more competition.

Competition for Cable? *Satellites provide consumers another way of accessing television broadcasts, not only in remote areas such as this ranch, but also in highly populated areas where the additional competition keeps cable prices down and quality up.*

Until 1992, the Federal Communications Commission (FCC), the federal agency that regulates the telecommunications industry, measured competition by the number of over-the-air channels. At first, the commission decided that three over-the-air channels represented effective competition. It did not regulate cable television companies in areas where there were more than three over-the-air channels. Later on, when it noticed that prices of cable television were rising and consumers were complaining, the FCC raised the limit to six over-the-air channels. In 1992, Congress passed a law saying that it did not matter how many channels there were; the law required the FCC to regulate cable television firms in any case.

Over-the-air channels are not the only competition for cable television. People can use satellite dishes, which provide access to numerous channels at a price competitive with cable. Eventually, it may be possible to use the telephone wires to transmit television signals, in which case the telephone companies could compete with the cable television companies.

High-tech industries change quickly, and it is difficult for government regulators to keep up with the changes. Inflexible regulatory rules could slow innovation. In fact, upon taking over as chairman of the FCC in 2005, some of the first statements Kevin Martin made were about his plans to loosen rules so that neither phone nor cable companies would be required to share their Internet connections with competitors like America Online. He argued that these legacy regulations discouraged companies from investing in high-speed Internet service.

Regulators as Captives of Industry

Government and government agencies are run by people who have their own motivations, such as being reelected or increasing their influence. Thus, despite the economic advice about what government regulatory agencies should do, the agencies may end up doing something else. In fact, regulators have sometimes ended up helping the industry at the expense of the consumer. The railroad industry is an example. Originally, regulation of railroads was set up to reduce prices below the monopoly price. But as competition to the railroads from trucks and eventually airlines increased, the industry continued to be regulated. Eventually, the regulators were helping the industry; they kept prices from falling to prevent railroad firms from failing. And by regulating trucking prices, they kept trucking firms from competing with the railroads. The Teamsters Union, which represents truck drivers, was one of the strongest supporters of regulation because it knew that the regulations were keeping trucking prices high. In a sense, the regulators became captives of both the firms and the workers in the industry.

An economist, George Stigler, won the Nobel Prize for showing how regulatory agencies could become captive to the industry and therefore tend to thwart competition. The concern that regulators will become captives is one reason some economists worry about allowing the government to regulate a new industry, like cable television. Eventually, the government may try to protect the cable television operators in order to prevent them from failing. The government might limit competition in the future from satellite dishes or from the telephone company.

The Deregulation Movement

deregulation movement: begun in the late 1970s, the drive to reduce the government regulations controlling prices and entry in many industries.

Starting in the late 1970s under Jimmy Carter and continuing in the 1980s under Ronald Reagan, the **deregulation movement**—the lifting of price regulations—radically changed several key industries. The list of initiatives that constitute this deregulation movement is impressive. For example, air cargo was deregulated in 1977, air travel was deregulated in 1978, satellite transmissions were deregulated in 1979, trucking was deregulated in 1980, cable television was deregulated in 1980 (although regulation was reimposed in 1992), crude oil prices and refined petroleum products were deregulated in 1981, and radio was deregulated in 1981. There was also deregulation of prices in the financial industry. Prior to the 1980s, the price—that is, the interest rate on deposits—was controlled by the financial regulators. Regulation of brokerage fees was also eliminated.

This deregulation of prices reduced deadweight loss. Airline prices have declined for many travelers. It is now cheaper to ship goods by truck or by rail. Economists have estimated the size of this reduction in deadweight loss by calculating the increase in the area between the demand curve and the marginal cost curve as the quantity produced increased.

Some people complain about deregulation. Business travelers complain that they have to pay more for air travel, although vacation travelers can pay less. Deregulation of the airline industry led to widespread fears that large airlines would dominate the industry because of their market power at the hubs. However, the large airlines are now so cost-heavy that smaller regional airlines have made significant headway in attracting even business travelers with their low-cost flights. Boston-based business travelers, for example, might be willing to suffer the inconvenience of traveling to Providence to take a cheaper Southwest Airlines flight rather than fly out of Boston on one of the large carriers.

REVIEW

- In many cases it is clear that a natural monopoly exists and thus price regulation is needed. However, in certain industries like cable television, there is controversy about the need for regulation.
- There has frequently been price regulation where there is no natural monopoly, as in trucking.
- The deregulation movement began in the late 1970s and continued into the 1980s. It was in response to economic analysis that showed that it is harmful to regulate the prices of firms that are not natural monopolies.
- Trucking, airline, and railroad transportation prices are lower as a result of this deregulation. As with most economic changes, not everyone benefited. Business travelers saw the costs of some services increase.

Conclusion

This chapter analyzed a key role of government in a market economy: maintaining competitive markets through antitrust policy or the regulation of firms. By reducing the deadweight loss due to monopoly, the government can reduce market failure and improve people's lives.

However, this analysis must be placed in the context of what in reality motivates government policymakers. The example of regulators becoming captives of industry reminds us that having an analysis of what should be done is very different from getting it done. Government failure is a problem that must be confronted just like market failure. Reducing government failure requires designing the institutions of government to give government decision-makers the proper incentives.

KEY POINTS

1. The government has an important role to play in maintaining competition in a market economy.
2. Part of antitrust policy is breaking apart firms with significant market power, although this technique is now used infrequently. Section 2 of the Sherman Antitrust Act provides the legal authority for challenging existing monopolies.
3. A more frequently used part of antitrust policy is preventing mergers that would cause significant market power. In the United States, the government must approve mergers.
4. Concentration measures such as the HHI are used to decide whether a merger should take place.
5. Price fixing is a serious antitrust offense in the United States, and the laws against it are enforced by allowing private firms to sue, providing for treble damages, and allowing the government to ask for criminal penalties.
6. In the case of natural monopolies, the government can either run the firm or regulate a private firm. In the United States, the latter route is usually taken.
7. Regulatory agencies have been using incentive regulation more frequently in order to give firms incentives to hold costs down.
8. The deregulation movement has consisted mainly of removing price regulations from firms that are not natural monopolies, such as trucking and airlines.
9. Overall the deregulation movement has significantly lowered costs to consumers, but it is controversial because services have been cut back in certain areas.

KEY TERMS

antitrust policy
Sherman Antitrust Act
rule of reason
predatory pricing
Clayton Antitrust Act
Federal Trade Commission (FTC)
Antitrust Division of the Justice Department
Herfindahl-Hirschman index (HHI)
market definition
horizontal merger
vertical merger
price fixing
treble damages
exclusive territories
exclusive dealing
resale price maintenance
marginal cost pricing
average total cost pricing
incentive regulation
deregulation movement

QUESTIONS FOR REVIEW

1. What historical development gave the impetus to the original antitrust legislation in the United States?
2. What is the difference between Section 1 and Section 2 of the Sherman Antitrust Act?
3. What is the difference between the rule of reason and the per se rule in the case of monopolization and in the case of price fixing?
4. What law gives the government the right to prevent mergers that would increase market power?
5. Why is the market definition crucial when calculating the HHI index in the case of mergers?
6. Why is marginal cost pricing a faulty pricing rule for regulatory agencies?
7. How does incentive regulation improve on average cost pricing?
8. Why is there more controversy about regulating cable television than about regulating water companies?

PROBLEMS

1. Which legislation—Section 1 of the Sherman Act, Section 2 of the Sherman Act, or the Clayton Act—gives the government the authority to take action in each of the following areas: prosecuting price fixing, preventing proposed mergers, breaking up existing monopolies, suing for predatory pricing?
2. Why is it better to break up monopolies that are not natural monopolies rather than regulate them, even if it is possible to regulate them?
3. In reflecting on a recent term of service, a former head of the Antitrust Division said, "I was convinced that a little bit of efficiency outweighs a whole lot of market power." Evaluate this statement by considering two sources of efficiency: decreasing average total cost and research and development. Describe how these should be balanced against the deadweight loss from market power.
4. Compare the following two hypothetical cases of price fixing.
   a. General Motors, Ford, and Chrysler are found to be coordinating their prices for Chevy Blazers, Ford Broncos, and Jeep Cherokees.

b. General Motors is coordinating with Chevy dealers around the country to set the price for Chevy Blazers.

Which is more likely to raise prices and cause a deadweight loss? Explain.

5. The following table shows the market shares of firms in three different industries.

Industry	Number of Firms	Shares	HHI
1	100	Each firm with 1 percent	
2	15	10 firms with 5 percent	
		5 firms with 10 percent	
3	3	1 firm with 60 percent	
		2 firms with 20 percent	

a. Complete the above table by calculating the Herfindahl-Hirschman index.

b. Will the FTC try to prevent a significant merger in industry 2? In industry 3? Why?

6. Use the merger guidelines to decide whether the following changes in industry C in Table 12.1 would be permitted.

a. The three small firms merge into one firm.

b. One of the firms with 4 percent share merges with the firm with 8 percent share.

7. If economies of scale are important, is it possible for consumers to be better off if the government allows more mergers?

8. In some states, regulatory authorities are beginning to allow some competition among electric power companies. What must the regulators think about the nature of this industry? What other industries have gone through this transformation? What are the benefits of deregulation?

9. Sketch a graph of a natural monopoly with declining average total cost and constant marginal cost.

a. Show how the monopoly causes a deadweight loss, with price not equal to marginal cost.

b. Describe the pros and cons of three alternative ways to regulate the monopoly and reduce deadweight loss: marginal cost pricing, average total cost pricing, and incentive regulation.

10. The demand schedule and total costs for a natural monopoly are given in the following table.

Price	Quantity	Total Costs
16	6	80
15	7	85
14	8	90
13	9	95
12	10	100
11	11	105
10	12	110
9	13	115
8	14	120
7	15	125
6	16	130
5	17	135
4	18	140

a. Why is this firm a natural monopoly? What will the monopoly price be? Calculate profits.

b. Suppose the government sees that this is a natural monopoly and decides to regulate it. If the regulators use average total cost pricing, what will the price and quantity be? What should profits be when the regulators are using average total cost pricing?

c. If the regulators use marginal cost pricing, what will the price and quantity be? Why is this policy difficult for regulators to pursue in practice? What are profits in this situation?

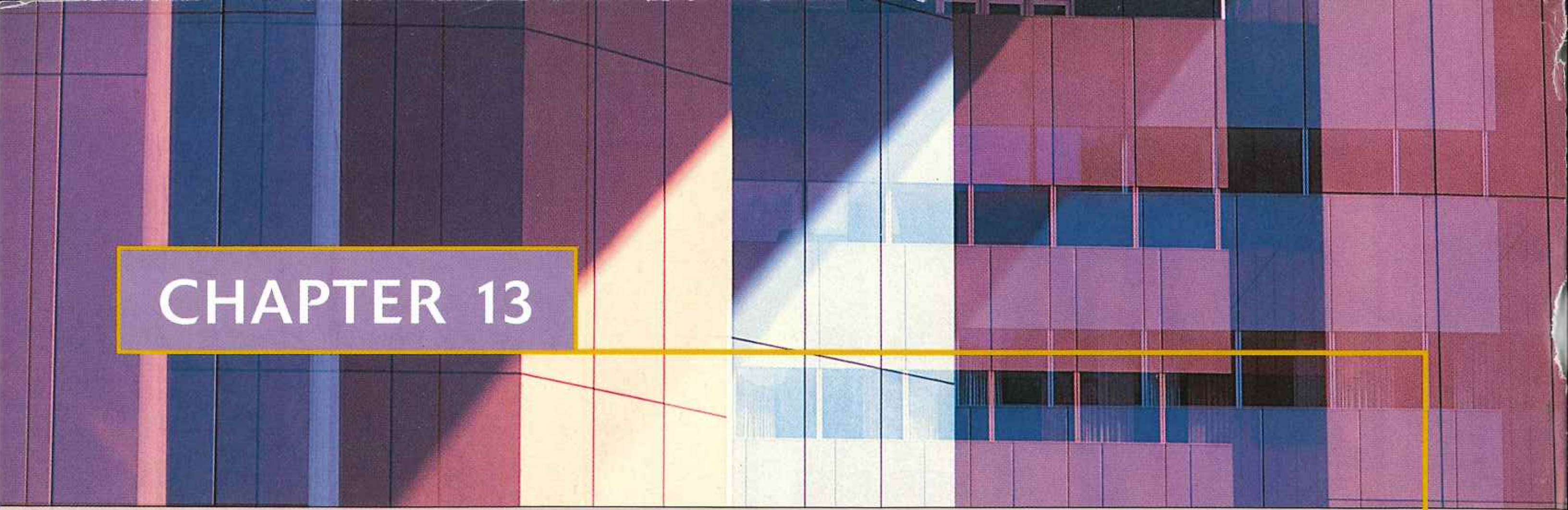
d. Why might the government want to use incentive regulation?

11. Historically, local telephone companies have been natural monopolies, but cellular phones are now offering services that are a substitute for wire connections.

a. If traditional phone companies are under incentive regulation, how would the introduction of cellular phones affect them? Show it graphically.

b. What should the regulatory agency do?

12. Some people argue that coal mines are natural monopolies. In fact, until recently, all coal mines in Great Britain were owned by the government. What conditions in the industry do you need to check in order to tell whether the industry is a natural monopoly?



CHAPTER 13

Labor Markets

What occupation will offer the best jobs in the future? What college major will provide the best chance of getting one of those jobs? Will women's earnings catch up to or exceed men's earnings in the future?

All these questions pertain to labor markets. Labor markets are the most pervasive markets in the world, touching many more people directly than stock markets do. For most people, income from the stock market is a small fraction of the wages and salaries earned in the labor market. It is not surprising, therefore, that many beginning economics students ask their teachers more questions about labor markets, such as the ones in the first paragraph, than they do about stock markets, even though many originally choose to take economics to learn more about the stock market.

In analyzing labor markets, economists stress their similarity to other markets; this enables economists to use the standard supply and demand model. To see the analogy, consider Figure 13.1, which illustrates a typical *labor market*. It shows a typical labor supply curve and typical labor demand curve. On the vertical axis is the price of labor, or the wage. On the horizontal axis is the quantity of labor, either the number of workers or the number of hours worked. People work at many different types of jobs—physical therapists, accountants, mechanics, teachers, Web developers, judges, professional athletes—and there is a labor market for each type. The labor market diagram in Figure 13.1 could refer to any one of these particular types of labor. The first thing to remember

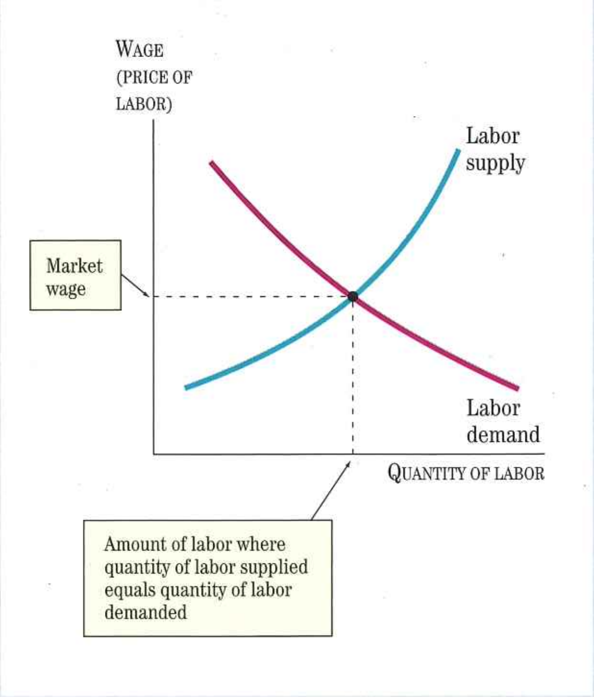


Figure 13.1
Labor Demand Curve and Labor Supply Curve
The basic economic approach to the labor market is to make an analogy with other markets. Labor is what is bought or sold on the labor market. The demand curve shows how much labor firms are willing to buy at a particular wage. The supply curve shows how much labor workers are willing to sell at a particular wage.

about the labor demand curve and the labor supply curve is that firms demand labor and people supply it. Labor—like other factors of production—is demanded by firms because it can be used to produce goods and services; the labor demand curve tells us the quantity of labor demanded by firms at each wage. The labor supply curve tells us the quantity of labor supplied by workers at each wage.

Note that the labor demand curve slopes downward and the labor supply curve slopes upward, just like other demand and supply curves. Thus, a higher wage reduces the quantity of labor demanded by firms, and a higher wage increases the quantity of labor supplied by people. Note also that the curves intersect at a particular wage and a particular quantity of labor. As with any other market, this intersection predicts the quantity of something (in this case, labor) and its price (in this case, the wage).

In this chapter, we show how the labor demand and supply model rests on the central economic idea that people make purposeful choices with limited resources and interact with other people when they make these choices. We will see that the model can be used to explain interesting facts about the labor market. We start by reviewing these facts, and we then explain why wages change over time and why there are gaps between the wages of skilled and unskilled workers, between the wages of women and men, and between the wages of union and nonunion workers. Even some of the problems caused by discrimination can be better understood using the standard tools of supply and demand.

Wage Trends

Are wages in the United States increasing? Are they increasing more rapidly or more slowly than in the recent past? In this section, we look at recent wage trends. First, we define exactly what is meant by the wage and show how it is measured.

Measuring Workers' Pay

When examining data on workers' pay, we must be specific about (1) what is included in the measure of pay, (2) whether inflation may be distorting the measure, and (3) the interval of time over which workers receive pay.

Pay Includes Fringe Benefits. Pay for work includes not only the direct payment to a worker—whether in the form of a paycheck, currency in a pay envelope, or a deposit in the worker's bank account—but also **fringe benefits.** Fringe benefits may consist of many different items: health or life insurance, when the employer buys part or all of the insurance for the employee; retirement benefits, where the employer puts aside funds for the employee's retirement; paid time off such as vacations and sick or maternity leave; and discounts on the company's products.

fringe benefits: compensation that a worker receives excluding direct money payments for time worked: insurance, retirement benefits, vacation time, and maternity and sick leave.

In recent years, fringe benefits have become an increasingly larger share of total compensation in the United States and many other countries. In the United States, fringe benefits are now about 29 percent of total pay. In 1960, fringe benefits were only about 8 percent of total pay.

The term *wage* sometimes refers to the part of the payment for work that excludes fringe benefits. For example, a minimum wage of $5.15 per hour does not usually include fringe benefits. But in most economics textbooks, the term **wage** refers to the *total* amount a firm pays workers, *including* fringe benefits. This book uses the usual textbook terminology. Thus, the wage is the price of labor.

wage: the price of labor defined over a period of time worked.

Adjusting for Inflation: Real Wages versus Nominal Wages. When comparing wages in different years, it is necessary to adjust for inflation, the general increase in prices over time. The **real wage** is a measure of the wage that has been adjusted for changes in inflation. The real wage is computed by dividing the stated wage by a measure of the price of the goods and services. The most commonly used measure for this purpose is the consumer price index (CPI), which gives the price of a fixed collection, or market basket, of goods and services each year compared to some base year. For example, the CPI increased from about 1.00 in the 1994 base year to 1.27 in 2004. This means that the same goods and services that cost $100 in 1983 cost $127 in 2004. Suppose the hourly wage for a truck driver increased from $10 to $19 from 1994 to 2004, or by 90 percent; then the real wage increased from $10 (= $10/1.00) to $14.96 (= $19/1.27), or an increase of about 50 percent. Thus, because of the increase in prices, the real wage gain for the truck driver was less than the 90 percent stated wage gain would suggest. The term *nominal wage* is used to emphasize that the wage has not been corrected for inflation. The real wage is the best way to compare wages in different years.

real wage: the wage or price of labor adjusted for inflation; in contrast, the nominal wage has not been adjusted for inflation.

The Time Interval: Hourly versus Weekly Measures of Pay. It is also important to distinguish between *hourly* and *weekly* measures of workers' pay. Weekly earnings are the total amount a worker earns during a week. Clearly, weekly earnings will be less for part-time work than for full-time work (usually 40 hours per week) if hourly earnings are the same.

Because part-time work has increased and the average number of hours per week has declined in the last 30 years in the United States, weekly earnings for the average worker have increased less rapidly than hourly earnings.

Wage Trends

Having described how to measure wage trends, let's now look at what has happened to wages in the United States in recent years. Figure 13.2 shows average real hourly wages in the United States. In 2004, the average wage was about $22.23 per hour, *including* a total of about $6.48 in fringe benefits.

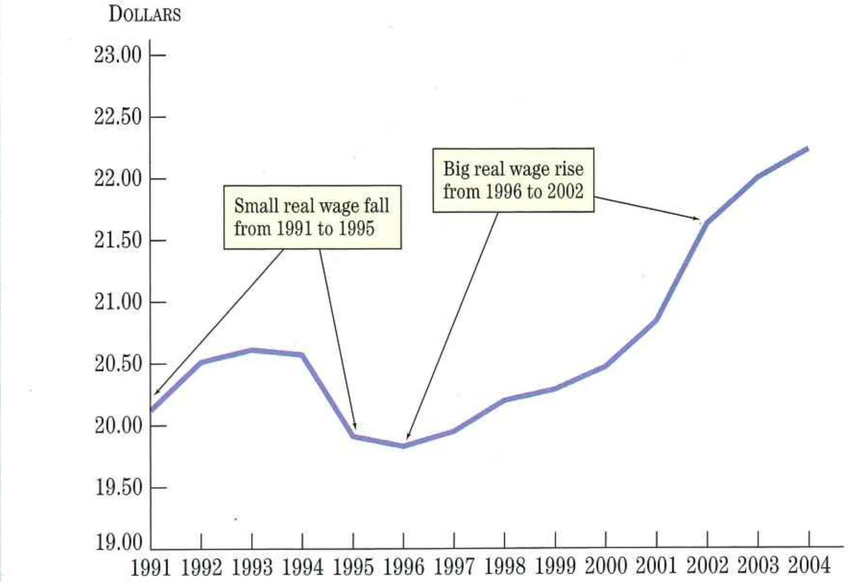


Figure 13.2
The Average Hourly Real Wage
In the United States, average real hourly wages (including fringe benefits) started to grow more rapidly in the mid-1990s. (Wages are in 1999 dollars.)

What is most noticeable in Figure 13.2 is that workers' pay began to rise more rapidly in the mid-1990s, after stagnating for several years. From 1996 to 2002, real wages rose by an average of 1.9 percent per year, a rate more than 6 times greater than the .3 percent average in the earlier years of the decade. Later in this chapter we will provide an economic explanation for this remarkable development and examine whether or not it will continue.

Figure 13.2 shows the average wage for all workers. What about the dispersion, or distribution, of wages across the population? Casual observation reveals large differences between the earnings of some people or groups and others. Sports celebrities and corporate executives are paid in the millions, many times the average wage in the United States. Workers with higher skills are paid more than workers with lower skills. College graduates earn more on average than those with a high school education or less. But there are other types of wage dispersions. For example, women on average earn less than men.

The distribution of wages across workers has also changed substantially in recent years. One development that has received much attention from economists is that the pay gap between skilled and less skilled workers has increased. In the mid-1970s, college graduates earned about 45 percent more than high school graduates. In the 1990s, this was up to about 65 percent.

Another change is in the wage difference between women and men, which, though still wide, has been narrowing in recent years. In the mid-1970s, women on average earned less than 60 cents for each dollar men earned. By the late 1990s, the gap had closed to around 76 cents.

What causes these changes? Can the economists' model of labor markets explain them? After developing the model in the next section, we will endeavor to answer these questions.

REVIEW

- Workers' pay includes fringe benefits, which are about 27 percent of workers' pay in the United States.
- The real wage is an adjustment of the nominal wage to account for changes in inflation.
- Real hourly wages have increased at a more rapid rate since the mid-1990s.
- The dispersion of wages between skilled and unskilled workers has increased, with college-educated workers gaining over those without a college education. The difference between the average wage of women and men is substantial, although it has narrowed in recent years.

Labor Demand

labor market: the market in which individuals supply their labor time to firms in exchange for wages and salaries.

labor demand: the relationship between the quantity of labor demanded by firms and the wage.

labor supply: the relationship between the quantity of labor supplied by individuals and the wage.

derived demand: demand for an input derived from the demand for the product produced with that input.

The **labor market** consists of firms that have a demand for labor and people who supply the labor. In this section, we look at **labor demand,** the relationship between the quantity of labor demanded by firms and the wage. In the next section, we look at **labor supply,** the relationship between the quantity of labor supplied by people and the wage. We start with a single firm's demand for labor and then sum up all the firms that are in the labor market to get the market demand for labor.

In deriving a firm's labor demand, economists assume that the firm's decision about how many workers to employ, like its decision about how much of a good or service to produce, is based on profit maximization. The demand for labor is a **derived demand;** that is, it is derived from the goods or services that the firm produces with the labor. The firm sells these goods and services to consumers in product markets, which are distinct from the labor market. Labor and other factors of production are not directly demanded by consumers; the goods and services labor produces are what is demanded by consumers. Thus, the demand for labor derives from these goods and services.

A Firm's Employment Decision

Recall how the idea of profit maximization was applied to a firm's decision about the quantity to produce: If producing another ton of steel will increase a steel firm's profits—that is, if the marginal revenue from producing a ton is greater than the marginal cost of producing the ton—then the firm will produce that ton of output. However, if producing another ton of steel reduces the firm's profits, then the firm will not produce that ton.

The idea of profit maximization is applied in a very similar way to a firm's decision about how many workers to employ: If employing another worker increases the firm's profits, then the firm will employ that worker. If employing another worker reduces the firm's profits, then the firm will not employ the worker.

We have already seen that a firm produces a quantity that equates marginal revenue to marginal cost (*MR* = *MC*). The firm satisfies an analogous condition in deciding how much labor to employ, as we discuss next.

■ **From Marginal Product to Marginal Revenue Product.** To determine a firm's demand curve for labor, we must examine how the firm uses labor to produce

Table 13.1
Labor Input and Marginal Revenue Product at a Competitive Firm

Workers Employed Each Week (*L*)	Quantity Produced (*Q*)	Marginal Product of Labor (*MP*)	Price of Output (dollars) (*P*)	Total Revenue (dollars) (*TR*)	Marginal Revenue Product of Labor (dollars) (*MRP*)
0	0	—	100	0	—
1	17	17	100	1,700	1,700
2	31	14	100	3,100	1,400
3	42	11	100	4,200	1,100
4	51	9	100	5,100	900
5	58	7	100	5,800	700
6	63	5	100	6,300	500
7	66	3	100	6,600	300
8	68	2	100	6,800	200
9	69	1	100	6,900	100
		$\frac{\text{Change in } Q}{\text{Change in } L}$	*P* does not depend on *Q*.	$P \times Q$	$\frac{\text{Change in } TR}{\text{Change in } L}$ or $P \times MP$

its output of goods and services. We start by assuming that the firm sells its output in a *competitive market*; that is, the firm is a *price-taker*. We also assume that the firm takes the wage as given in the labor market; in other words, the firm is hiring such a small proportion of the workers in the labor market that it cannot affect the market wage for those workers. Table 13.1 gives an example of such a competitive firm. It shows the weekly production and labor input of a firm called Getajob, which produces professional-looking job résumés in a college town. To produce a résumé, workers at Getajob talk to each of their clients—usually college seniors—give advice on what should go into the résumé, and then produce the résumé.

The first two columns of Table 13.1 show how Getajob can increase its production of résumés each week by employing more workers. This is the *production function* for the firm; it assumes that the firm has a certain amount of capital—word-processing equipment, a small office near the campus, and so on. We assume that labor is the only variable input to production in the short run, so that the cost of increasing the production of résumés depends only on the additional cost of employing more workers. Observe that the *marginal product (MP) of labor*—which we defined in Chapter 6 as the change in the quantity produced when one additional unit of labor is employed—declines as more workers are employed. In other words, there is a diminishing marginal product of labor, or diminishing return to labor: As more workers are hired with office space and equipment fixed, each additional worker adds less and less to production. For example, the first worker employed can produce 17 résumés a week, but if there are already 8 workers at Getajob, hiring a ninth worker will increase production by only 1 résumé.

marginal product of labor: the change in production due to a one-unit increase in labor input. (Ch. 6)

Suppose that the market price for producing this type of résumé service is $100 per résumé, as shown in the fourth column of Table 13.1. Because Getajob is assumed to be a *competitive firm*, it cannot affect this price. Then, the total revenue of the firm for each amount of labor employed can be computed by multiplying the price (P) times the quantity produced (Q) with each amount of labor (L). This is shown in the next-to-last column. For example, total revenue with $L = 3$ workers employed is $P = \$100$ times $Q = 42$, or $4,200.

marginal revenue product of labor: the change in total revenue due to a one-unit increase in labor input.

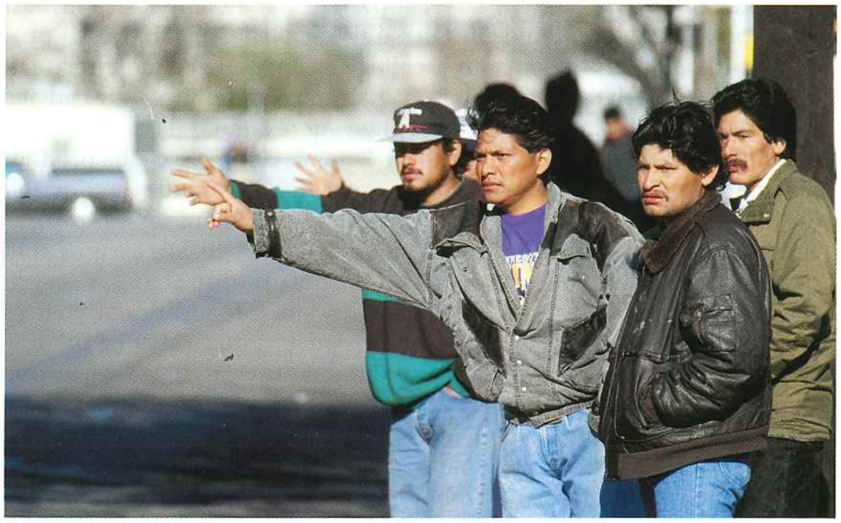
The last column of Table 13.1 shows the **marginal revenue product (*MRP*) of labor.** *The marginal revenue product of labor is defined as the change in total revenue when one additional unit of labor is employed.* For example, the marginal revenue product of labor from hiring a third worker is the total revenue with 3 workers ($4,200) minus the total revenue with 2 workers ($3,100), or $4,200 − $3,100 = $1,100. The marginal revenue product of labor is used to find the demand curve for labor, as we will soon see.

Wait. Before you read any further, make sure you can explain the difference between marginal product (*MP*) and marginal revenue product (*MRP*).

What is the difference between the marginal product (*MP*) and the marginal revenue product (*MRP*)? The marginal product is the increase in the *quantity produced* when labor is increased by one unit. The marginal revenue product is the increase in *total revenue* when labor is increased by one unit. For a *competitive firm* taking the market price as given, the marginal revenue product (*MRP*) can be calculated by multiplying the marginal product (*MP*) by the price of output (*P*). For example, the marginal product when the third worker is hired is 11 résumés; thus, the additional revenue that the third worker will generate for the firm is $100 per résumé times 11, or $1,100.

Observe in Table 13.1 that the marginal revenue product of labor declines as more workers are employed. This is because the marginal product of labor declines.

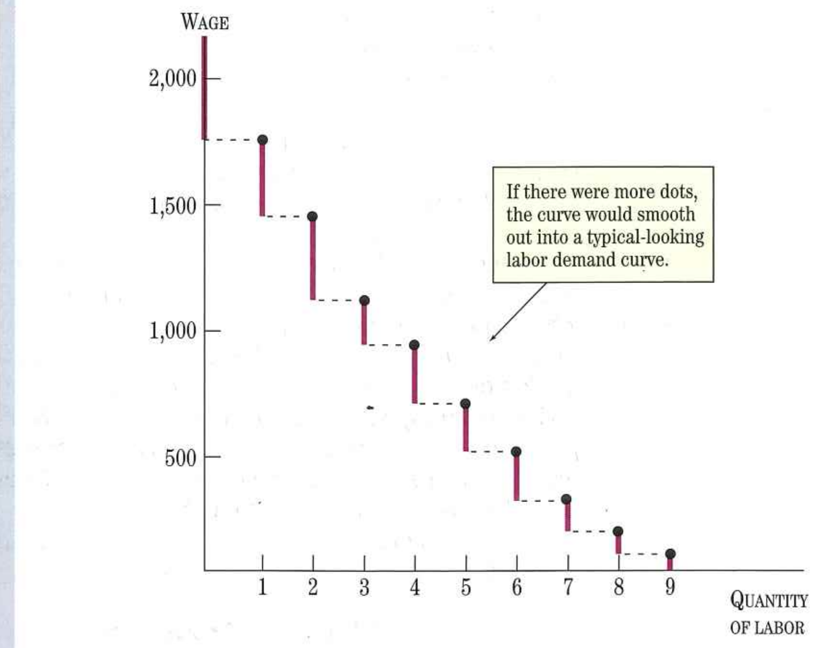
The Marginal Revenue Product of Labor Equals the Wage ($MRP = W$). Now we are almost ready to derive the firm's demand curve for labor. Suppose first that the wage for workers with the type of skills Getajob needs in order to produce résumés is $600 per week (for example, $15 per hour for 40 hours). Then, hiring 1 worker certainly makes sense because the marginal revenue product of labor is $1,700, or much greater than the $600 wage cost of hiring the worker. How about 2 workers? The marginal revenue product from employing a second worker is $1,400, still greater than $600, so it makes sense to hire a second worker. Continuing this way, we see that the *firm will hire a total of 5 workers when the wage is $600 per week,* because hiring a sixth worker would result in a marginal revenue product of only $500, less than the $600 per week wage.



Looking for Work
These day workers are job hunting for construction clean-up work at a downtown street corner in Austin, Texas. The workers signal their availability by showing up at the street corner; labor contractors will come by and hire the number of workers they need that day.

Thus, if a firm maximizes profits, it will hire the largest number of workers for which the marginal revenue product of labor is greater than the wage; if fractional units of labor input (for example, hours rather than weeks of work) are possible, then the firm will keep hiring workers until the marginal revenue product of labor exactly equals the wage. Thus, we have derived a key rule of profit maximization: Firms will hire workers up to the point where the *marginal revenue product of labor equals the wage.*

The rule that the marginal revenue product of labor equals the wage can be written in symbols as $MRP = W$.



If there were more dots, the curve would smooth out into a typical-looking labor demand curve.

Figure 13.3
Determining a Firm's Demand Curve for Labor
The black dots are exactly the same as the marginal revenue product of labor in Table 13.1. The red line indicates the quantity of labor demanded at each wage.

The Firm's Derived Demand for Labor

Now, to find the demand curve for labor, we need to determine how many workers the firm will hire at *different* wages. We know that Getajob will hire 5 workers if the wage is $600 per week. What if the wage is $800 per week? Then the firm will hire only 4 workers; the marginal revenue product of the fifth worker ($700) is now less than the wage ($800), so the firm will not be maximizing its profits if it hires 5 workers. Thus we have shown that a higher wage reduces the quantity of labor demanded by the firm. What if the wage is lower than $600? Suppose the wage is $250 a week, for example. Then the firm will hire 7 workers. Thus, a lower wage increases the quantity of labor demanded by the firm.

Figure 13.3 shows how to determine the entire demand curve for labor. It shows the wage on the vertical axis and the quantity of labor on the horizontal axis. The plotted points are the marginal revenue products from Table 13.1. To find the demand curve, we ask how much labor the firm would employ at each wage. Starting with a high wage, we reduce the wage gradually, asking at each wage how much labor the firm would employ. At a weekly wage of $2,000, the marginal revenue product is less than the wage, so it does not make sense to hire any workers. Therefore, the quantity demanded is zero at wages above $2,000. At a weekly wage of $1,500, it makes sense to hire one worker, and so on. As the wage is gradually lowered, the quantity of labor demanded rises, as shown by the red line in Figure 13.3. The step-like downward-sloping curve is the labor demand curve. There would be more black dots and the curve would be very smooth if we measured work in fractions of a week rather than in whole weeks.

Table 13.2
Labor Input and Marginal Revenue Product for a Firm with Power to Affect the Market Price

Workers Employed Each Week (*L*)	Quantity Produced (*Q*)	Marginal Product of Labor (dollars) (*MP*)	Price of Output (dollars) (*P*)	Total Revenue (dollars) (*TR*)	Marginal Revenue Product of Labor (dollars) (*MRP*)
0	0	—	100	0	—
1	17	17	92	1,564	1,564
2	31	14	85	2,635	1,071
3	42	11	79	3,318	683
4	51	9	75	3,825	507
5	58	7	71	4,118	293
6	63	5	69	4,347	229
7	66	3	67	4,422	75
8	68	2	66	4,488	66
9	69	1	65	4,485	−3

$\frac{\text{Change in } Q}{\text{Change in } L}$

P declines with *Q*.

P × *Q*

$\frac{\text{Change in } TR}{\text{Change in } L}$

Observe in Figure 13.3 that a firm's demand curve for labor is completely determined by the firm's marginal revenue product of labor. We have shown why the demand curve for labor is downward-sloping: because the marginal revenue product of labor curve is downward-sloping. A higher wage will reduce the quantity of labor demanded, and a lower wage will increase the quantity of labor demanded; these are *movements along* the downward-sloping labor demand curve. We also can explain why a firm's labor demand curve would *shift*. For example, if the price (*P*) of the good (résumés) rises—perhaps because the demand curve for résumés shifts outward—then the marginal revenue product of labor ($MRP = P \times MP$) will rise and the demand curve for labor will shift outward. Similarly, a rise in the marginal product of labor (*MP*) will shift the labor demand curve outward. On the other hand, a decline in the price (*P*) or a decline in the marginal product (*MP*) will shift the labor demand curve to the left.

What If the Firm Has Market Power? This approach to deriving the demand curve for labor works equally well for the case of a firm that is not a price-taker but is instead a monopoly or a monopolistic competitor. Table 13.2 shows an example of such a firm. The key difference between the firm in Table 13.1 and the firm in Table 13.2 is in the column for the price. Rather than facing a constant price for its output and thus a horizontal demand curve, this firm faces a downward-sloping demand curve: It can increase the quantity of résumés demanded by lowering its price. For example, if Getajob's résumés are slightly differentiated from those of other résumé producers in town, then the demand curve that Getajob faces when selling résumés may be downward-sloping.

Once we observe that the price and output are inversely related, we can continue just as we did with the competitive firm. Again, total revenue is equal to the price times the quantity, and marginal revenue product is the change in total revenue as 1 more worker is hired. Again, the marginal revenue product declines as more workers are hired, as shown in the last column of Table 13.2. However, now the marginal

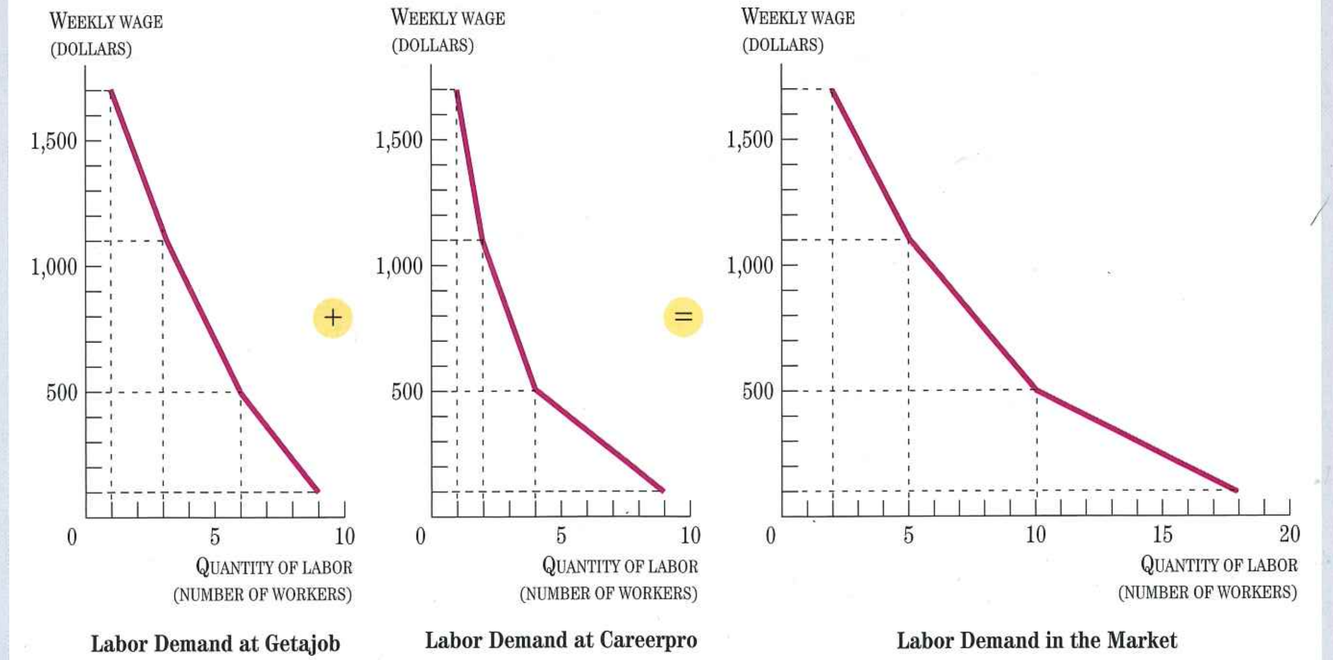


Figure 13.4 Summing Firms' Demands to Get the Labor Market Demand Curve

The labor demand curve in the market is obtained by summing the quantities of labor demanded by all the firms at each wage.

revenue product declines more sharply as more workers are employed, and it even turns negative. The reason is that as more workers are hired and more output is produced and sold, the price of output must fall. This cuts into revenue, even though output increases, because the lower price on items previously sold reduces revenue. But the principle of labor demand is the same: Firms hire up to the point where the marginal revenue product of labor equals the wage. The marginal revenue product determines the labor demand curve.

In the case of a firm with market power, the simple relationship $MRP = P \times MP$ no longer holds, however, because the firm does not take the market price as given. Instead, we replace the price (P) by the more general marginal revenue (MR) in that relationship. This implies that the marginal revenue product is equal to the marginal revenue (MR) times the marginal product (MP). The relationship $MRP = MR \times MP$ holds for all firms, whether they have market power or not. Only for a competitive firm is $MR = P$.

■ **Market Demand for Labor.** To get the demand for labor in the market as a whole, we must add up the labor demand curves for all the firms demanding workers in the labor market. At each wage, we sum the total quantity of labor demanded by all firms in the market; this is illustrated in Figure 13.4 for the case of two firms producing résumés. The two curves on the left are labor demand curves for two résumé-producing firms, Getajob and Careerpro. (The curves are smoothed out compared with Figure 13.3 so that they are easier to see.) The process of summing individual

Table 13.3
Marginal Cost and the Production Decision at Getajob

Workers Employed Each Week (*L*)	Quantity Produced (*Q*)	Variable Costs (dollars) (*VC*)	Marginal Cost (dollars) (*MC*)
0	0	0	0
1	17	600	35
2	31	1,200	43
3	42	1,800	55
4	51	2,400	67
5	58	3,000	86
6	63	3,600	120
7	66	4,200	200
8	68	4,800	300
9	69	5,400	600

$600 wage × *L*

$\frac{\text{Change in } VC}{\text{Change in } Q}$

firms' demands for labor to get the market demand is analogous to summing individual demand curves for goods to get the market demand curve for goods. At each wage, we sum the labor demand at all the firms to get the market demand.

A Comparison of *MRP* = *W* with *MC* = *P*

Note that a firm's decision to employ workers is closely tied to its decision about how much to produce. We have emphasized the former decision here and the latter decision in earlier chapters. To draw attention to this connection, we show in Table 13.3 the marginal cost when the wage is $600. Marginal cost is equal to the change in variable costs divided by the change in quantity produced. Variable costs are the wage times the amount of labor employed.

Now, consider the quantity of output the firm would produce if it compared price and marginal cost as discussed in earlier chapters. If the price of output is $100, the firm will produce 58 résumés, the highest level of output for which price is greater than marginal cost. This is exactly what we found using the *MRP* = *W* rule, because 58 units of output requires 5 workers. Recall that employing 5 workers is the profit-maximizing labor choice when the wage is $600.

If the profit-maximizing firm could produce fractional units, then it would set marginal cost exactly equal to price (*MC* = *P*). The resulting production decision would be exactly the same as that implied by the rule that the marginal revenue product of labor equals the wage.

REVIEW

- The demand for labor is a relationship between the quantity of labor a firm will employ and the wage.
- The demand for labor is a derived demand because it is derived from the goods and services produced by labor. When the quantity of labor is the decision variable, the firm maximizes profits by setting the marginal revenue product of labor equal to the wage.

- When the wage rises, the quantity of labor demanded by firms declines. When the wage falls, the quantity of labor demanded increases. These are movements along the labor demand curve.
- When the price of a commodity produced by a particular type of labor rises, the demand curve for that type of labor shifts outward.

Labor Supply

We now focus on *labor supply*. The market labor supply curve is the sum of many people's individual labor supply curves. The decision about whether to work and how much to work depends very much on individual circumstances.

Work versus Two Alternatives: Home Work and Leisure

Consider a person deciding how much to work—either how many hours a week or how many weeks a year. As with any economic decision, we need to consider the alternative to work. Economists have traditionally called the alternative *leisure*, although many of the activities that make up the alternative to work are not normally thought of as leisure. These activities include "home work," like painting the house or caring for children at home, as well as pure leisure time, such as simply talking to friends on the telephone, going bowling, or hiking in the country. The price of leisure is the opportunity cost of not working, that is, the wage. If a person's marginal benefit from more leisure is greater than the wage, then the person will choose more leisure. The decision to consume more leisure is thus like the decision to consume more of any other good. This may seem strange, but the analogy works quite well in practice.

Effects of Wage Changes: Income and Substitution Effects. Like the decision to consume a commodity, the decision to work can be analyzed with the concepts of the *substitution effect* and the *income effect*.

The *substitution effect* says that the higher the wage, the more attractive work will seem compared to its alternatives: home work or leisure. A higher wage makes work more rewarding compared to the alternatives. Think about your own work opportunities. You may have many nonwork choices, including studying, sleeping, and watching TV. Although you enjoy these activities, suppose that the wage paid for part-time student employment triples. Then you might decide to work an extra hour each day. The sacrifice—less time to study, sleep, watch TV, and so on—will be worth the higher wage. The inducement to work a little more because of the higher wage is the substitution effect. The quantity of labor supplied tends to increase when the wage rises because of the substitution effect.

The *income effect*, as in the demand for goods, reflects the effect of the price change on your real income. For example, if the wage for student employment triples, and you are already working, you might think that you can work less. With a higher wage, you can earn the same amount by working less. You may even have more money to do things other than work. Note that the income effect works in the opposite direction from the substitution effect: The quantity of labor supplied tends to decrease, rather than increase, when the wage rises because of the income effect.

The Shape of Supply Curves. Because the substitution effect and the income effect work in opposite directions, the labor supply curve can slope either upward or downward. The supply curve slopes upward if the substitution effect dominates but slopes downward if the income effect dominates. Several possibilities for labor supply curves are illustrated in Figure 13.5.

Moreover, the same supply curve may slope upward for some range of wages and downward for another range. For example, at high wage levels—when people earn enough to take long vacations—the income effect may dominate. At lower wages, the substitution effect may be dominant. This would then result in a **backward-bending labor supply curve,** as shown in Figure 13.6.

backward-bending labor supply curve: the situation in which the income effect outweighs the substitution effect of an increase in the wage at higher levels of income, causing the labor supply curve to bend back and take on a negative slope.

This derivation of the labor supply curve may seem unrealistic. After all, the workweek is 40 hours for many jobs; you may not have much choice about the number of hours per week. In fact, the sensitivity of the quantity of labor supplied to the wage is probably small for many workers. But economists have shown that the effect is large for some workers, and therefore it is useful to distinguish one worker's supply curve from another's.

In a family with two adults and children, for example, one of the adults may already have a job and the other may be choosing between working at home and working outside the home. This decision may be very sensitive to the wage and perhaps the cost of child care or of consuming more prepared meals. In fact, the increased number of women working outside the home may be due to the increased opportunities and wages for women. The increase in the wage induces workers to work more in the labor market. Economists have observed a fairly strong wage effect on the amount women work, as illustrated in the Reading the News About box on page 345.

One also needs to distinguish between the effects of a temporary change in the wage and a more permanent change. Empirical studies show that the quantity of labor supplied rises more in response to a temporary increase in the wage than to

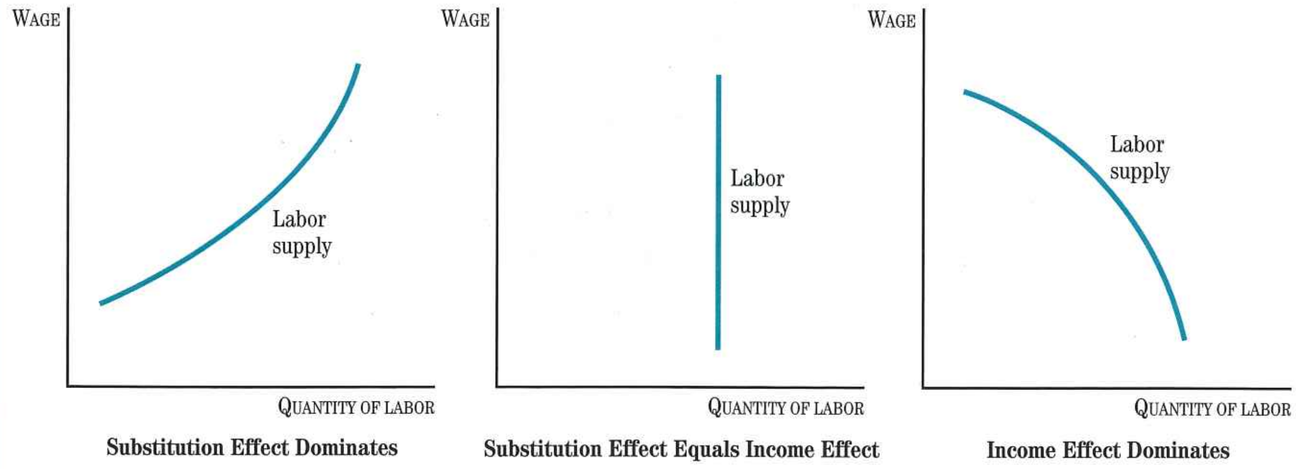
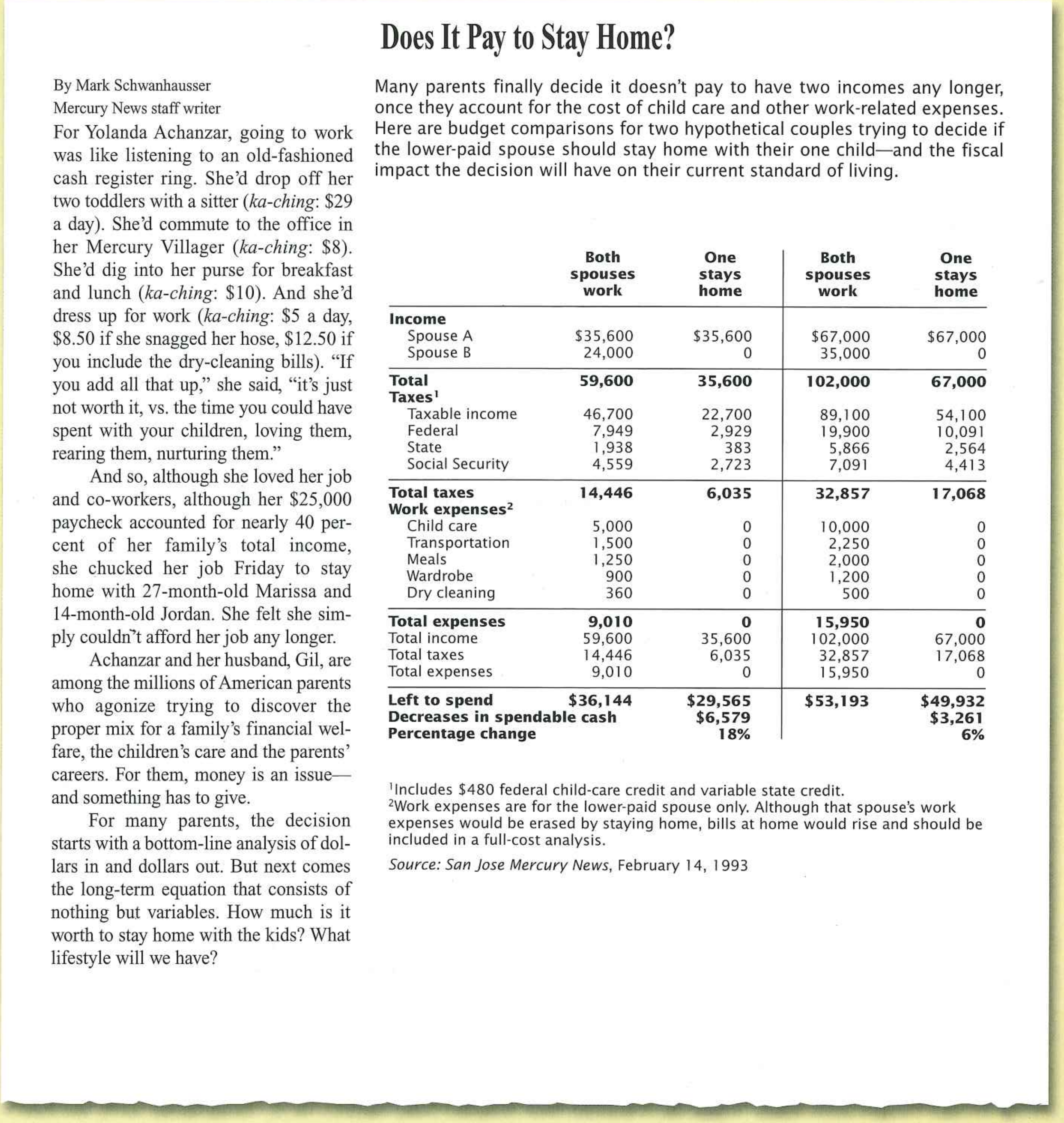


Figure 13.5
Three Labor Supply Curves

The three curves differ in the relative strength of the income and substitution effects. The labor supply curve on the left slopes upward because the substitution effect is stronger than the income effect. For the curve on the right, the income effect is stronger than the substitution effect. For the vertical curve in the middle, the two effects are the same.

READING THE NEWS ABOUT... Incentives to Work

As more and more families have two potential workers, the decision about labor supply has become a household decision. The following newspaper article from the *San Jose Mercury News* (February 14, 1993) tells a story that pairs the human side of the decision with the economic side. According to the calculations in the table, the net earnings from work—after taxes and all other expenses—may be very small in some cases.



Does It Pay to Stay Home?

By Mark Schwanhausser
Mercury News staff writer

For Yolanda Achanzar, going to work was like listening to an old-fashioned cash register ring. She'd drop off her two toddlers with a sitter (*ka-ching*: $29 a day). She'd commute to the office in her Mercury Villager (*ka-ching*: $8). She'd dig into her purse for breakfast and lunch (*ka-ching*: $10). And she'd dress up for work (*ka-ching*: $5 a day, $8.50 if she snagged her hose, $12.50 if you include the dry-cleaning bills). "If you add all that up," she said, "it's just not worth it, vs. the time you could have spent with your children, loving them, rearing them, nurturing them."

And so, although she loved her job and co-workers, although her $25,000 paycheck accounted for nearly 40 percent of her family's total income, she chucked her job Friday to stay home with 27-month-old Marissa and 14-month-old Jordan. She felt she simply couldn't afford her job any longer.

Achanzar and her husband, Gil, are among the millions of American parents who agonize trying to discover the proper mix for a family's financial welfare, the children's care and the parents' careers. For them, money is an issue—and something has to give.

For many parents, the decision starts with a bottom-line analysis of dollars in and dollars out. But next comes the long-term equation that consists of nothing but variables. How much is it worth to stay home with the kids? What lifestyle will we have?

Many parents finally decide it doesn't pay to have two incomes any longer, once they account for the cost of child care and other work-related expenses. Here are budget comparisons for two hypothetical couples trying to decide if the lower-paid spouse should stay home with their one child—and the fiscal impact the decision will have on their current standard of living.

	Both spouses work	One stays home	Both spouses work	One stays home
Income				
Spouse A	$35,600	$35,600	$67,000	$67,000
Spouse B	24,000	0	35,000	0
Total	**59,600**	**35,600**	**102,000**	**67,000**
Taxes[1]				
Taxable income	46,700	22,700	89,100	54,100
Federal	7,949	2,929	19,900	10,091
State	1,938	383	5,866	2,564
Social Security	4,559	2,723	7,091	4,413
Total taxes	**14,446**	**6,035**	**32,857**	**17,068**
Work expenses[2]				
Child care	5,000	0	10,000	0
Transportation	1,500	0	2,250	0
Meals	1,250	0	2,000	0
Wardrobe	900	0	1,200	0
Dry cleaning	360	0	500	0
Total expenses	**9,010**	**0**	**15,950**	**0**
Total income	59,600	35,600	102,000	67,000
Total taxes	14,446	6,035	32,857	17,068
Total expenses	9,010	0	15,950	0
Left to spend	**$36,144**	**$29,565**	**$53,193**	**$49,932**
Decreases in spendable cash		**$6,579**		**$3,261**
Percentage change		**18%**		**6%**

[1] Includes $480 federal child-care credit and variable state credit.

[2] Work expenses are for the lower-paid spouse only. Although that spouse's work expenses would be erased by staying home, bills at home would rise and should be included in a full-cost analysis.

Source: San Jose Mercury News, February 14, 1993

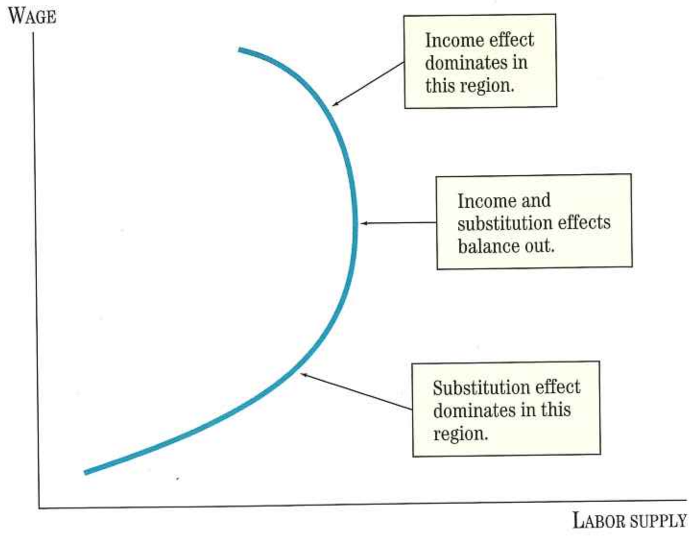


Figure 13.6
Backward-Bending Labor Supply Curve
A person may have a labor supply curve that is positively sloped for a low wage, is steeper for a higher wage, and then bends backward for a still higher wage.

a permanent increase. What's the explanation? Consider an example. If you have a special one-time opportunity tomorrow to earn $100 an hour rather than your usual $6 an hour, you are likely to put off some leisure for one day; the substitution effect dominates. But if you are lucky enough to land a lifetime job at $100 an hour rather than $6 an hour, you may decide to work fewer hours and have more leisure time; the income effect dominates.

This difference between temporary and permanent changes helps explain the dramatic decline in the average hours worked per week in the United States as wages have risen over the last century. These are more permanent changes, for which the income effect dominates.

Work versus Another Alternative: Getting Human Capital

The skills of a worker depend in part on how much schooling and training the worker has. The decision to obtain these skills—to finish high school and attend a community college or obtain a four-year college degree—is much like the choice between work and leisure. In fact, an important decision for many young people is whether to go to work or to finish high school; if they have finished high school, the choice is whether to go to work or to go to college.

Economists view the education and training that raise skills and productivity as a form of *investment*, a decision to spend funds or time on something now because it pays off in the future. Continuing the analogy, an investment in a college education raises the amount of **human capital**—a person's knowledge and skills—in the same way that the investment in a factory or machine by a business firm raises physical capital. Figure 13.7 demonstrates the kind of difference this investment can make.

human capital: a person's accumulated knowledge and skills.

The decision to invest in human capital can be approached like any other economic choice. Suppose the decision is whether Angela should go to college or get a

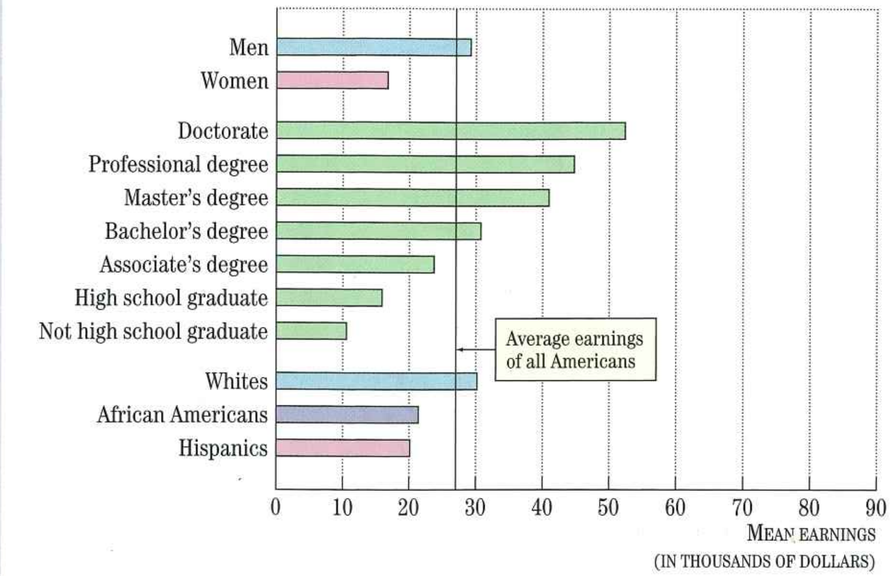


Figure 13.7
Higher Education and Economic Success
According to this chart, education pays off in terms of earnings, with doctorate degree holders earning the most, followed by workers with professional and master's degrees.

job. If she does not go to college, she saves on tuition and can begin earning an income right away. If she goes to college, she pays tuition and forgoes the opportunity to earn income at a full-time job. However, if Angela is like most people, college will improve her skills and land her a better job at higher pay. The returns on college education are the extra pay. Angela ought to go to college—invest in human capital—if the returns are greater than the cost.

on-the-job training: the building of the skills of a firm's employees while they work for the firm.

People can increase their skills at work as well as in school. In fact, **on-the-job training** is one of the most important ways in which workers' productivity increases. On-the-job training can be either *firm-specific,* where the skills are useful only at one firm, or *general-purpose,* where the skills are transferable to other jobs.

REVIEW

- The labor supply curve can be viewed as the outcome of an individual's choice between work and some other activity, whether home work, schooling, or leisure.
- There is both a substitution effect and an income effect on the labor supply. The substitution effect is the increased attractiveness of work relative to its alternative as the wage rises. The income effect is the increased attractiveness of leisure because there is more to spend when the wage rises. In some situations the income effect dominates. In other situations the substitution effect dominates.
- Human capital is the knowledge and skills that a person accumulates from going to school and receiving on-the-job training. The return on human capital has increased in recent years.

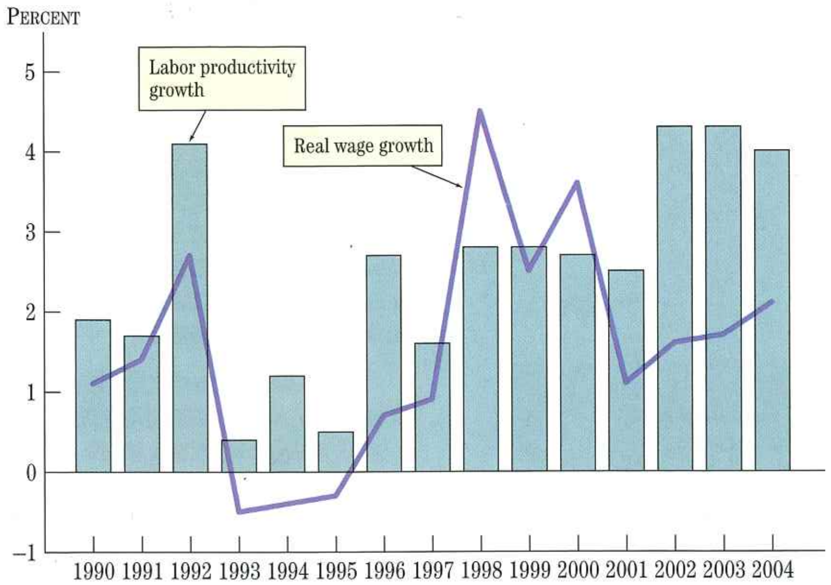


Figure 13.8
Labor Productivity Growth and Real Wage Growth
Labor productivity growth is closely related to the growth of real wages, much as would be predicted by the labor supply and demand model.

Explaining Wage Differences

When we combine the labor demand and labor supply curves derived in the previous two sections, we get the model of the labor market summarized in Figure 13.1. The model predicts that the wage in the labor market will be at the intersection of the supply and demand curves. The point of intersection, where the quantity of labor supplied equals the quantity of labor demanded, is the **labor market equilibrium.**

labor market equilibrium: the situation in which the quantity of labor supplied equals the quantity of labor demanded.

Labor Productivity

The model of the labor market predicts that the wage equals the marginal revenue product. If the marginal product of labor employed at a firm increases, then the model predicts that the wage will rise. Suppose the marginal product of labor rises for the economy as a whole; then wages should also rise. Is this what occurs in reality?

The Wage Boom and Labor Productivity Boom. In Figure 13.8, the line graph shows the percentage by which real wages have increased each year since 1990, using the wage data we examined earlier in this chapter. Note that wages rose rapidly starting in 1996. The bars in Figure 13.8 show output per hour of work in the same period. Output per hour of work is called **labor productivity** and is a good indication of trends in the marginal product of labor on average in the United States. The labor market model predicts that wages in the United States should increase when labor productivity increases. Do they?

labor productivity: output per hour of work.

Figure 13.8 shows a strong correlation between labor productivity and the real wage. Note that the change in the labor productivity trend occurred in the mid-

Economics in Action

Does Productivity or Compensating Differentials Explain the Academic Wage Gap?

People with Ph.D.'s who teach or do research at colleges and universities are paid 10 percent less than people with Ph.D.'s who work for government and 20 percent less than people with Ph.D.'s who work for business firms. Why does the academic wage gap exist?

There are two possibilities: (1) People with Ph.D.'s who work in business and government are more skilled and more productive, or (2) people with Ph.D.'s who work in business and government are paid a compensating wage differential because the job is less pleasant. (They don't have the pleasure of teaching students or the flexible academic hours.)

How can we tell which is the right explanation? Looking at what happens to people with Ph.D.'s when they move provides an answer. If their wages increase when they move to a nonacademic job, then compensating wage differentials rather than productivity differences is the correct explanation.

The following table shows the average salary increases between 1985 and 1987 of people with Ph.D.'s who either (1) did not move, (2) moved to another college or university, or (3) moved to business or government. The salary increases are largest for those who left academia. For example, the average salary increase for engineering Ph.D.'s nearly doubled when they moved from academia to work in a business firm or government. This indicates that the differences are due not to skill but to compensating wage differentials.

This is one of the rare cases in which economists have actually been able to obtain data that distinguish compensating wage differentials from productivity or other explanations for wage differences. But if the case is representative, compensating wage differentials may play a big part in wage dispersion.

Largest increase for every type of Ph.D.

	Increase in Salary (dollars)		
	Did Not Move	*Moved to Another College*	*Left Academia*
Physical science	7,303	10,216	15,330
Mathematical science	6,523	9,716	15,727
Environmental science	6,292	4,688	11,333
Life science	5,870	6,710	8,115
Psychology	5,920	6,559	10,371
Social science	5,796	7,687	12,485
Engineering	7,294	6,724	14,025
Humanities	5,042	5,380	8,204

Source: Adapted from Albert Rees, "The Salaries of Ph.D.'s in Academia and Elsewhere," *Journal of Economic Perspectives*, Winter 1993. Reprinted with permission.

1990s, at almost the same time as the change in the trend of real wage growth. The close empirical association between wages and labor productivity that is evident in this chart suggests that labor productivity is a key explanation of wage changes over time.

■ **Wage Dispersion and Productivity.** Can labor productivity also explain wage differences between people? If the marginal product of labor increases with additional skills from investment in human capital, then, on average, wages for people with a college education should be higher than those for people without a college education. Hence, productivity differences are an explanation for the wage gap between workers who do not receive education beyond high school and those who are college educated.

Although human capital differences undoubtedly explain some of the dispersion of wages, some people have argued that the greater productivity of college-educated workers is due not to the skills learned in college but to the fact that colleges screen applicants. For example, people who are not highly motivated or who have difficulty communicating have trouble getting into college. Hence, college graduates would earn higher wages even if they learned nothing in college. If this is so, a college degree *signals* to employers that the graduate is likely to be a productive worker.

Unfortunately, it is difficult to distinguish the skill-enhancing from the signaling effects of college. Certainly your grades and your major in college affect the kind of job you get and how much you earn, suggesting that more than signaling is important to employers. In reality, signaling and human capital both probably have a role to play in explaining the higher wages of college graduates.

Whether it is signaling or human capital that explains the higher wages of college graduates, labor productivity differences are still the underlying explanation for the wage differences. However, labor productivity does not explain everything about wages. Consider now some other factors.

Compensating Wage Differentials

Not all jobs that require workers with the same level of skill and productivity are alike. Some jobs are more pleasant, less stressful, or safer than other jobs. For example, the skills necessary to be a deep-sea salvage diver and a lifeguard are similar—good at swimming, good judgment, and good health. But the risks—such as decompression sickness—for a deep-sea diver are greater and the opportunity for social interaction is less. If the pay for both jobs were the same, say, $10 per hour, most people would prefer the lifeguard job.

High Wages for High Work
Compensating wage differentials are illustrated by the relatively high wages paid to someone for performing risky jobs such as window washing on a skyscraper.

But this situation could not last. With many lifeguard applicants, the beach authorities would be unlikely to raise the wage above $10 and might even try to cut the wage if budget cuts occurred. With few applicants, the deep-sea salvage companies would have to raise the wage. After a while, it would not be surprising to see the wage for lifeguards at $9 per hour and the wage for deep-sea divers at $12 per hour; these wages would be labor market equilibrium wages in the supply and demand model for lifeguards and deep-sea divers. Thus, we would be in a situation where the skills of the workers were identical but their wages were much different. The higher-risk job pays a higher wage than the lower-risk job.

Situations in which wages differ because of the characteristics of the job are widespread. Hazardous duty pay is common in the military. Wage data show that night-shift workers in manufacturing plants are paid wages that are about 3 percent higher on average than those of daytime workers, presumably to compensate for the inconvenience.

compensating wage differential: a difference in wages for people with similar skills based on some characteristic of the job, such as riskiness, discomfort, or convenience of the time schedule.

Such differences in wages are called **compensating wage differentials.** They are an important source of differences in wages that are not based on marginal product. With compensating differentials, workers may seek out riskier jobs in order to be paid more.

Discrimination

As noted earlier, the gap in earnings between women and men has been narrowing in recent years. Women now make close to 80 percent of the wages of men, whereas 50 years ago, women earned only about 50 percent of the wages of men. The gap is also closing for blacks and whites, although not quite as quickly. In the 1950s, the ratio of wages of blacks to that of whites was about 60 percent; it has narrowed to about 70 percent since then. Wage differences between white and minority workers and between men and women are an indication of discrimination if the wage differences cannot be explained by differences in marginal product or other factors unrelated to race or gender.

Wage Differences for Workers with the Same Marginal Products. Some, but not all, of these differences may be attributed to differences in human capital. The wage gaps between blacks and whites and between men and women with comparable education and job experience are smaller than the ratios in the preceding paragraph. But a gap still exists.

Discrimination on the basis of race or gender prejudice can explain such differences. This is shown in Figure 13.9. *Discrimination* can be defined in the supply and demand model as not hiring women or minority workers even though their marginal product is just as high as that of other workers, or paying a lower wage to such workers even though their marginal product is equal to that of other workers. Either way, discrimination can be interpreted as a leftward shift of the labor demand curve for women or minority workers. As shown in Figure 13.9, this reduces the wages and employment for those discriminated against.

Competitive Markets and Discrimination. An important implication of this supply and demand interpretation of the effects of discrimination is that competition among firms may reduce it. This is an advantage of competitive markets that should be added to the advantages already mentioned. Why might competition reduce discrimination? Because firms in competitive markets that discriminate will lose out to firms that do not. Much like firms that do not keep their costs as low as other firms, they will eventually be driven out of the industry.

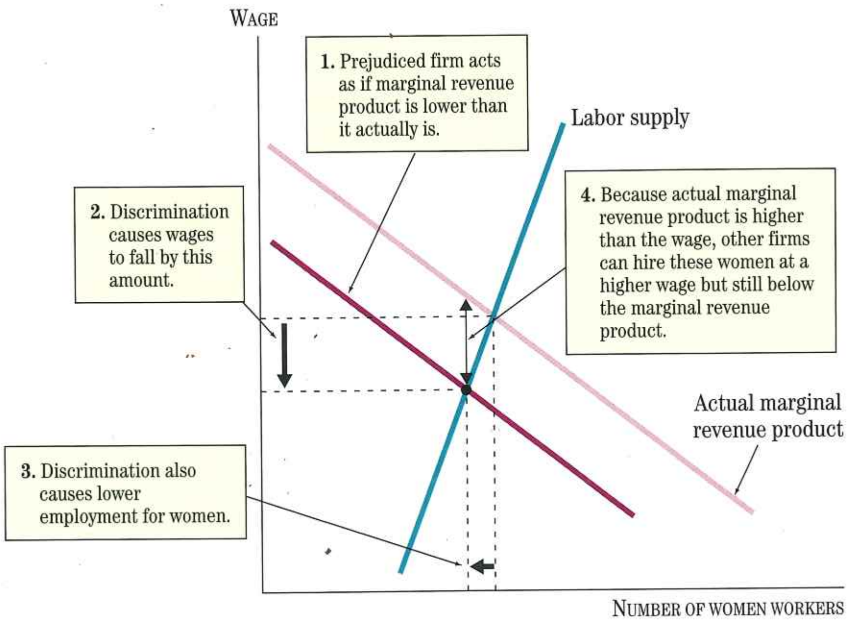


Figure 13.9
Discrimination in the Labor Market
Firms that discriminate against women pay them a wage that is less than their marginal product. But this gives other firms an opportunity to recruit workers from prejudiced firms by paying higher wages.

If markets are competitive, then firms that discriminate against women or minorities will pay them a wage lower than their marginal revenue product, as shown in Figure 13.9. In this situation, any profit-maximizing firm will see that it can raise its profits by paying these workers a little more—but still less than their marginal revenue product—and hiring them away from firms that discriminate. As long as the discriminating firms pay less than the marginal product of labor, other firms can hire the workers and raise profits. Remember that a firm will increase profits if the wage is less than the marginal revenue product. But eventually competition for workers will raise wages until the wages are equal to the marginal products of labor.

This description of events relies on a market's being competitive. If firms have monopoly power or entry is limited, so that economic profits are not driven to zero, then discrimination can continue to exist. That discrimination effects on wages do persist may be a sign that there is market power and barriers to entry. In any case, there are laws against discrimination that give those who are discriminated against for race, gender, or other reasons the right to sue those who are discriminating.

Some laws have been proposed requiring that employers pay the same wage to workers with comparable skills. Such proposals are called *comparable worth proposals.* The intent of such proposals is to bring the wages of different groups into line. However, such laws might force wages to be the same in situations where wages are different for reasons other than discrimination, such as compensating wage differentials. This would lead to shortages or surpluses, much as price ceilings or price floors in any market do. In the lifeguard/deep-sea diver example, a law requiring employers to pay lifeguards and deep-sea divers the same wage would cause a surplus of lifeguards and a shortage of deep-sea divers. For example, suppose that with comparable worth legislation, the wage for both lifeguards and deep-sea divers was $10 per hour. Because the labor market equilibrium wage for lifeguards, $9 per hour, is less

THE FAR SIDE® By GARY LARSON
Larson
HELP WANTED
SEPT 2, 2005
NINTENDO EXPERT NEEDED $50,000 salary + bonus Equal opportunity employer 555-
LOOKING FOR GOOD MARIO BROTHERS PLAYER $100,000 plus your own car. 555-
CAN YOU SAVE THE PRINCESS? we need skilled men & women $75,000 + Retirement.
Expanding Company needs skilled computer games operator. Call 571
IF YOU HAVE 50,000 HOURS OR MORE OF VIDEO GAME EXPERIENCE, WE NEED YOU!
DO YOU LAUGH IN THE FACE OF KILLER GOOMBAS? CALL US. $80,000 yr. plus a free house.
SUPER MARIO BROS. Expert $95,000 yr. Four-day work week + Ferrari
DO YOU KNOW A NINTENDO EXPERT? Please read him or her this ad
VIDEO PLAY dragon slay $150,000
NINTEN OFFE Sala
© 1990 FarWorks, Inc. All Rights Reserved/Dist. by Creators Syndicate

Hopeful parents

The Far Side® by Gary Larson © 1990 FarWorks, Inc. All Rights Reserved. Used with permission.

than $10 per hour, there would be a surplus of lifeguards: More people would be willing to be lifeguards than employers would be willing to hire. And because the labor market equilibrium wage for deep-sea divers, $12 per hour, is greater than $10, there would be a shortage of deep-sea divers: Firms would be willing to hire more deep-sea divers than the number of deep-sea divers willing to dive for the $10 per hour wage.

Minimum Wage Laws

Another example in which the government stipulates a wage that employers must pay is *minimum wage legislation,* which is common in many countries. The minimum wage sets a floor for the price of labor. Because wages differ due to skills, the impact of the minimum wage depends on the skills of the workers. Figure 13.10 shows what the supply and demand model predicts about the impact of the minimum wage on skilled and unskilled workers. A labor market for unskilled workers is shown on the left; the minimum wage is shown to be above the labor market equilibrium wage. There is thus a surplus, or unemployment: The quantity of labor demanded by firms at the minimum wage is less than the quantity of labor workers are willing to supply at that wage. A labor market for skilled workers is shown on the right: The minimum wage is shown to be below the market equilibrium wage for skilled workers. Thus

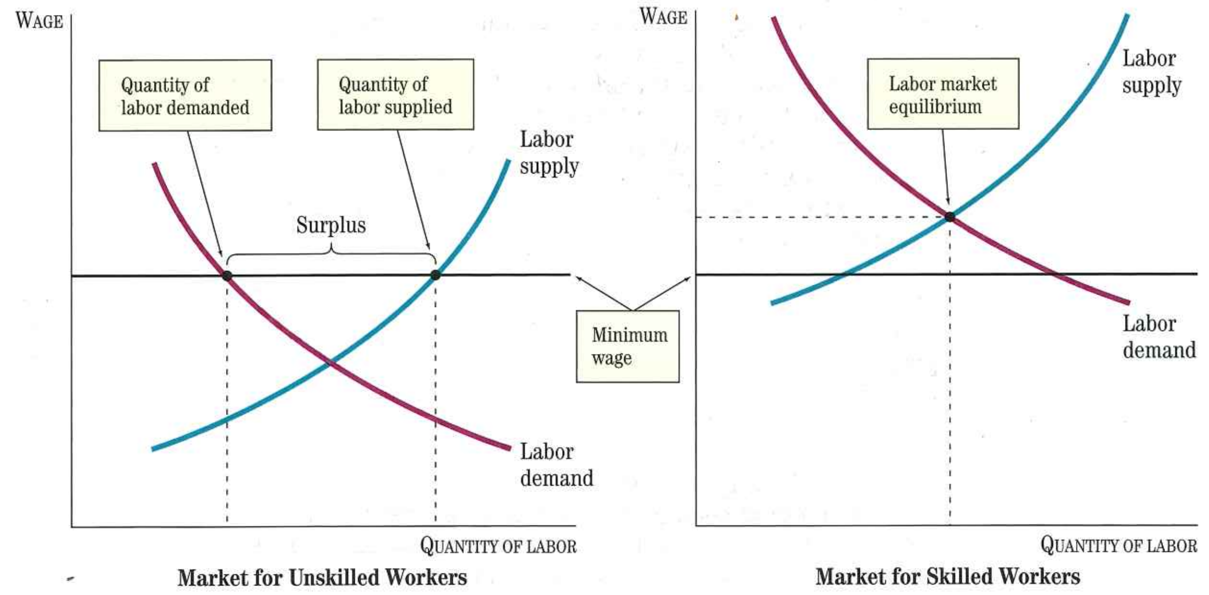


Figure 13.10
Effects of a Minimum Wage

If there are government restrictions on wages that hold them above the labor market equilibrium, then a surplus (unemployment) will arise. Unemployment is more likely for unskilled workers because the equilibrium wage is lower.

a minimum wage at the level shown in the graph would not cause unemployment among skilled workers.

Therefore, the labor supply and demand model predicts that the minimum wage is a cause of unemployment among less skilled or less experienced workers, and thereby ends up hurting some of the least well off in society. This is why many economists are concerned about the impact of minimum wage legislation.

In interpreting this result, remember that the supply and demand model is a *model* of reality, not reality itself. Although the model explains much about wages, its predictions about minimum wage laws should be verified like the predictions of any other economic model. In fact, labor economists have been trying to check the predictions of the model for the minimum wage for many years. Some, such as Jacob Mincer of Columbia University, have provided evidence that the predicted minimum wage impact is verified in real-world labor markets. Others, such as David Card of the University of California at Berkeley and Alan Krueger of Princeton University, have examined the effects of different minimum wage laws on different states on low-skilled fast-food workers and have not found evidence of the predicted impact on unemployment. David Neumark of Michigan State University and William Wascher of the Federal Reserve have disputed Card's and Krueger's data and found that the minimum wage enacted in those same states did have a negative effect on employment. Because of this controversy, testing the supply and demand model of labor has been a hot topic for the past twenty years.

REVIEW

- Labor productivity differences are an explanation for some of the differences in wages.
- Compensating wage differentials occur because some jobs are more attractive than others. They are another source of wage disparity.
- Discrimination reduces the wages of those who are discriminated against below their marginal revenue product.
- Data suggest that the wage effects of discrimination continue to exist but have declined in recent years. The female–male wage gap and the black–white wage gap have declined but are still significant.
- Competition can be a force against the effects of discrimination.

Wage Payments and Incentives

The agreement to buy or sell labor is frequently a long-term one. Job-specific training and the difficulty of changing jobs make quick turnover costly for both firms and workers. Thus employers and workers need to have an understanding of what will happen in the future, when, for example, the marginal product of labor at the firm increases or decreases.

Most workers would prefer a certain wage to an uncertain one; such workers will prefer a fixed wage that does not change every time the marginal revenue product changes. Long-term arrangements of this kind are quite common. A worker—a person working at Getajob, for example—is hired at a given weekly wage. If marginal revenue product declines because of a week of stormy winter weather with frequent power outages, Getajob will not reduce the weekly wage. On the other hand, when a crowd of college seniors arrives at the shop in May, the Getajob workers will have to work harder—their marginal revenue product will rise—but they will not be paid a higher wage. Thus, the weekly wage does not change with the actual week-to-week changes in the marginal revenue product of the worker. The wage reflects marginal revenue product over a longer period. Most workers in the United States are paid in this way.

Piece-Rate Wages

piece-rate system: a system by which workers are paid a specific amount per unit they produce.

An alternative wage payment arrangement endeavors to match productivity with the wage much more closely. Such contracts are used when the weekly or hourly wage does not provide sufficient *incentive* or where the manager cannot observe the worker carefully. Under a **piece-rate system,** the specific amount workers are paid depends on how much they produce. Thus, if their marginal product drops off, for whatever reason, they are paid less. Piece rates are common in the apparel and agriculture industries.

Consider California lettuce growers, for example. The growers hire crews of workers to cut and pack the lettuce. A typical crew consists of two cutters and one packer, who split their earnings equally. The crew is paid a piece rate, about $1.20 for a box of lettuce that might contain two dozen heads. A three-person crew can pick and pack about 75 boxes an hour. Thus, each worker can earn about $30 an hour. But if they slack off, their wages decline rapidly.

On the same lettuce farms, the growers may pay other workers on an hourly or weekly basis. The workers who wash the lettuce are paid an hourly wage. Truck drivers and the workers who carry the boxes to the trucks are also paid by the hour.

Why the difference? Piece rates are used when incentives are important and it is difficult to monitor the workers. This would apply to small crews of lettuce workers out in the fields but not to workers washing lettuce at the main building. Another reason is that some jobs, like washing lettuce or driving a truck, require particular care and safety. Workers might drive the truck too fast or wash the lettuce carelessly under a piece-rate system.

Deferred Wage Payments

deferred payment contract: an agreement between a worker and an employer whereby the worker is paid less than the marginal revenue product when young, and subsequently paid more than the marginal revenue product when old.

Yet another payment arrangement occurs when a firm pays workers less than their marginal revenue product when they are young and more than their marginal revenue product at a later time as a reward for working hard. Lawyers and accountants frequently work hard at their firms when they are young; if they do well, they make partner and are then paid much more than their marginal revenue product when they are older. Such contracts are called **deferred payment contracts.**

Generous retirement plans are another form of deferred payment contract. A reward for staying at the firm and working hard is a nice retirement package.

REVIEW

- Many labor market transactions are long term.
- Most employees receive a fixed hourly or weekly wage, even though their marginal revenue product fluctuates.
- Piece-rate contracts adjust the payment directly according to actual marginal product; they are a way to increase incentives to be more productive.
- Deferred compensation is another form of payment that aims at improving incentives and worker motivation.

Labor Unions

labor union: a coalition of workers, organized to improve the wages and working conditions of the members.

industrial union: a union organized within a given industry, whose members come from a variety of occupations.

craft union: a union organized to represent a single occupation, whose members come from a variety of industries.

The model of labor supply and demand can also help us understand the impact of labor unions. **Labor unions** such as the United Auto Workers or the United Farm Workers are organizations with the stated aim of improving the wages and working conditions of their members. There are two types of unions: **Industrial unions** represent most of the workers in an industry—such as the rubber workers, farm workers, or steelworkers—regardless of their occupation; **craft unions** represent workers in a single occupation or group of occupations, such as printers or dockworkers. In the 1930s and 1940s, there were disputes between those organizing craft unions and industrial unions. John L. Lewis, a labor union leader, argued that craft unions were not suitable for large numbers of unskilled workers. Hence, he and other union leaders split in 1936 from the American Federation of Labor (AFL), a group representing many labor unions, and formed the Congress of Industrial Organizations (CIO). It was not until 1955 that the AFL and CIO resolved their disputes and merged; one of the reasons for their resolution was that union membership was beginning to decline.

But the decline continued. In 2005, there was a split within AFL-CIO. Three large unions representing service workers, truck drivers, and food and commercial workers withdrew from the AFL-CIO expressing their unhappiness over the decline of union membership. About 12.5 percent of the U.S. labor force is currently unionized, down from about 25 percent in the mid-1950s. The fraction is much higher in other countries.

Unions negotiate with firms on behalf of their members in a collective bargaining process. Federal law, including the National Labor Relations Act (1935), gives workers the right to organize into unions and bargain with employers. The National Labor Relations Board has been set up to make sure that firms do not illegally prevent workers from organizing and to monitor union elections of leaders.

In studying unions, it is important to distinguish between the union leaders who speak for the union members and the union members themselves. Like politicians, union leaders must be elected, and as with politicians, we can sometimes better understand the actions of union leaders by assuming that they are motivated by the desire to be elected or reelected.

The Collective Voice of Union Workers
The lockout of dock workers in West Coast ports in the fall of 2002 paralyzed billions of dollars' worth of cargo going in and out of the United States. At the heart of the conflict was the introduction of new technology that would eliminate 200 to 600 jobs.

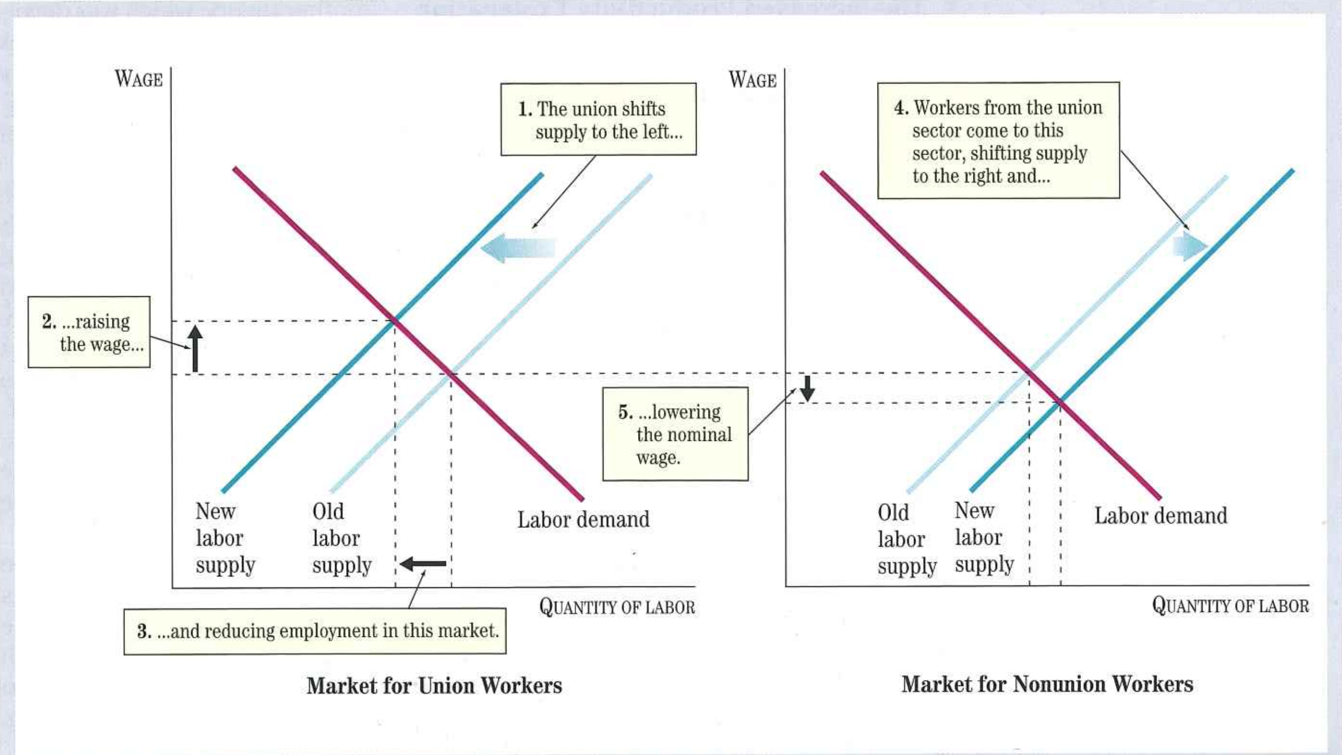

WAGE
1. The union shifts supply to the left...
2. ...raising the wage...
New labor supply
Old labor supply
Labor demand
3. ...and reducing employment in this market.
QUANTITY OF LABOR
Market for Union Workers
WAGE
4. Workers from the union sector come to this sector, shifting supply to the right and...
5. ...lowering the nominal wage.
Old labor supply
New labor supply
Labor demand
QUANTITY OF LABOR
Market for Nonunion Workers

Figure 13.11 The Effect of Unions on Wages

According to one view of labor unions, the union wage is increased relative to nonunion wages as a result of restricting the supply of workers. The supply in the union sector is reduced, but the supply in the nonunion sector is increased.

Union/Nonunion Wage Differentials

Studies of the wages of union workers and nonunion workers have shown that union wages are about 15 percent higher than nonunion wages, even when workers' skills are the same. There are two different explanations of how unions raise wages.

The Restricted Supply Explanation. One theory is that unions raise wages by restricting supply. By restricting membership, for example, they shift the labor supply curve to the left, raising wages, just as a monopolist raises the price of the good it sells by restricting supply. But when a union restricts supply, workers outside the union in another industry get paid less.

This effect of unions is illustrated in Figure 13.11. The graph on the right is one industry; the graph on the left is another industry. Suppose both industries require workers of the same skill level. Imagine the situation before the union is formed. Then the wages for the workers on the left and on the right in Figure 13.11 would be the same.

Now suppose a union organizes the industry on the left. Wages rise in the industry on the left, but the quantity of labor demanded in the industry falls. The workers in the industry on the left who become unemployed will probably move to the industry on the right. As they do so, the labor supply curve in the right-hand graph of Figure 13.11 shifts and the wage in that industry declines. Thus, a wage gap between the similarly skilled union and nonunion workers is created.

The Increased Productivity Explanation. Another theory, which was developed extensively in the book *What Do Unions Do?* by Richard Freeman and James Medoff of Harvard University, is that labor unions raise the wages of workers by increasing their marginal product. They do this by providing a channel of communication with management, motivating workers, and providing a democratic means of resolving disputes.

A worker who has a dispute with the management of a firm or who sees the opportunity to get a higher wage at another firm could, in principle, move. But such moves can have huge costs: The firm may have invested in job-specific training and the worker might like the area where the firm is located. In situations where exit from a firm is costly, people find other ways to improve their situation without exiting. The economist Albert Hirschman, in a famous book called *Exit, Voice, and Loyalty*, has called this alternative "voice." Rather than exit or quit, the worker may try to show the firm that a raise is deserved. Or the worker can discuss with the employer how conditions can be changed. The choice between exit and voice arises in many contexts: Should you transfer to a new college or tell the dean how the teaching might be improved? Should parents send their children to a private school or work to improve the local public school?

In many situations, exercising your voice requires collective action. If you alone complain to the dean, nothing much will happen, but if you organize a "students against lousy teaching" group, you may see some changes. Those who emphasize this collective-voice role of labor unions argue that unions provide a means through which workers improve their productivity. This explains why the wages of union workers are higher than those of nonunion workers with the same skills and training.

Monopsony and Bilateral Monopoly

The analysis of labor unions in Figure 13.11 stresses the market power of unions as *sellers* of their members' labor in the labor market: By restricting supply, the union can raise the price of its members' wages, much as a monopolist or a group of oligopolists with market power can raise the price of the goods they sell.

However, the *buyers* in the labor market—that is, the firms that purchase the labor—may also have market power to affect the wage, contrary to the assumption we have made throughout this chapter that firms do not have such market power in the labor market. **Monopsony** is a situation in which there is only one buyer. By reducing its demand, a monopsony can reduce the price in the market; it moves down along the supply curve, with both quantity and price lower.

monopsony: a situation in which there is a single buyer of a particular good or service in a given market.

bilateral monopoly: the situation in which there is one buyer and one seller in a market.

The situation in which there is only one seller (a monopoly) and one buyer (a monopsony) in a market is called a **bilateral monopoly.** A labor market with one labor union deciding the labor supply and one firm deciding the labor demand is an example of a bilateral monopoly.

In fact, there are few examples of monopsony; for most types of workers—sales clerks, accountants, engineers—there are typically many potential employers. Exceptions are found in small towns, where, for example, there may be only one auto repair shop. Then, if auto mechanics do not want to move, the auto repair shop is effectively the only employer. Another exception is found in professional sports leagues, where team owners form agreements with one another restricting workers' (that is, the players') mobility between teams. Such restrictions have been loosened significantly in recent years but still exist. If players had more freedom to move between teams, the teams' monopsony power would be reduced. Indeed, the loosening of restrictions that has already occurred has led to huge increases in players' salaries.

The outcome of a bilateral monopoly is difficult to predict. Compared to a situation where a monopsony faces competitive sellers, however, the bilateral monopoly can lead to a more efficient outcome. A firm with a monopsony facing many competitive sellers would buy *less* than a group of competitive buyers, in order to drive down the wage. By banding together, the sellers can confront this monopsony power with their own monopoly power. For example, they could refuse to work for less than the competitive wage. If their refusal is credible, they could take away the incentive for the monopsony to reduce labor demand because doing so would not reduce the wage.

REVIEW

- About 14 percent of U.S. workers belong to either industrial or craft unions.
- Workers who belong to unions are paid about 15 percent more on average than workers with the same skills who are not in unions. There are two conflicting explanations about why.
- One explanation is that labor unions improve productivity by improving worker motivation and providing workers with a collective voice.
- Another view is that labor unions raise productivity by restricting supply, much as a monopolist would, rather than by increasing productivity.

Conclusion and Some Advice

In this chapter, we have shown that the labor supply and demand model is a powerful tool with many applications. In fact, the model may apply to you, so consider carefully what it implies.

First, increasing your own labor productivity is a good way to increase your earnings. Many of the large differences in wages across individuals and across time are due to differences in productivity. Productivity is enhanced by increases in human capital, whether obtained in school or on the job. Such human capital will also prove useful if your firm shuts down and you need to find another job.

Second, if you are choosing between two occupations that you like equally well, choose the one that is less popular with other students of your generation and for which it looks like demand will be increasing. Both the supply and the demand for labor affect the wage, and if the supply is expected to grow more rapidly than the demand in the occupation you are training for, wages will not be as high as in the occupation for which labor is in relatively short supply.

Third, be sure to think about the wage you receive or the raises you get in real terms, not nominal terms, and make sure you are aware of fringe benefits offered or not offered.

Fourth, think about your job in a longer-term perspective. Partly for incentive reasons, some jobs pay little at the start, with the promise of higher payments later.

KEY POINTS

1. Wage growth in the United States, which is defined by the real hourly average pay (including fringe benefits), has been increasing at a faster rate since the mid-1990s. Wage dispersion has also increased.
2. The demand for labor is a derived demand that comes from the profit-maximizing decisions of firms. Firms adjust their employment to make the marginal revenue product of labor equal to the wage. For a competitive firm, the marginal product equals the wage divided by the price.
3. The supply curve for labor can be explained by looking at the choices of individuals or households. A person will work more hours if the wage is greater than the marginal benefit of more leisure.

4. The substitution effect and the income effect work in opposite directions, so that the labor supply curve can be either upward-sloping, vertical, downward-sloping, or backward-bending.
5. Long-term movements in wages are closely correlated with changes in labor productivity. Labor productivity differences also explain some of the differences in wages paid to different people.
6. Productivity does not explain everything. Compensating wage differentials and discrimination are other reasons wages differ.
7. When worker incentives or motivation are a problem, piece rates and deferred compensation can be used as alternative forms of payment arrangements.
8. Union workers earn more than nonunion workers who have the same skills. This occurs either because unions increase labor productivity or because they restrict the supply of workers in an industry.

KEY TERMS

fringe benefits
wage
real wage
labor market
labor demand
labor supply
derived demand
marginal revenue product of labor
backward-bending labor supply curve
human capital
on-the-job training
labor market equilibrium
labor productivity
compensating wage differential
piece-rate system
deferred payment contract
labor union
industrial union
craft union
monopsony
bilateral monopoly

QUESTIONS FOR REVIEW

1. Are fringe benefits a significant part of average pay in the United States?
2. Why is labor demand a derived demand?
3. What is the marginal revenue product, and why must it equal the wage if a firm is maximizing profits?
4. Why is the demand for labor downward-sloping?
5. Why do the substitution effect and the income effect on labor supply work in opposite directions?
6. How can compensating wage differentials explain why workers with the same skills are paid different amounts?
7. Why does discrimination against women and minorities reduce their wage, and why does competition reduce the effects of discrimination?
8. Why are piece rates sometimes used instead of weekly wages?
9. What is the difference between the two main views of labor unions?

PROBLEMS

1. Marcelo farms corn on 500 acres in a competitive industry, receiving $3 per bushel. The relationship between the number of workers Marcelo hires and production of corn is shown in the next column.

Number of Workers	Corn Production (bushels per year)
1	30,000
2	43,000
3	51,000
4	55,000
5	57,000
6	58,000

   a. Calculate the marginal product and marginal revenue product of labor for Marcelo's farm.
   b. If the wage for farm workers is $8,000 per year, how many workers will Marcelo hire? Explain.
   c. Suppose the yearly wage for farm workers is $8,000, the fixed rent is $30,000 per year, and there are no other costs. Calculate Marcelo's profits or losses. Will there be entry or exit from this industry?
2. Real wages in the United States are higher than in Guatemala. Using the supply and demand model for labor, explain why a difference in marginal product might explain this. Name one factor that may cause this difference in marginal product and explain its effect.
3. Draw a typical supply and demand for labor diagram to represent the market for doctors. Suppose a government regulation does not allow the wage rate for this profession to go as high as the market-determined wage rate. Depict this in your diagram. Will there be a shortage or surplus of doctors at that wage rate?

4. Use the definition of the demand for labor as the marginal revenue product to argue that the increasing wage dispersion between skilled and unskilled workers could come from (1) increases in the relative productivity of skilled workers and (2) increases in the demand for the products produced by skilled workers.
5. Given your answer to problem 4, what policies can the government pursue to correct this wage dispersion by affecting labor demand? What kinds of policies would the government pursue if it wanted to affect the supply of labor to correct excessive wage dispersion?
6. College professors are frequently paid less than others with equivalent skills working outside academia. Use the idea of compensating differentials to explain why professors' wages are relatively low.
7. A toy manufacturing company is considering hiring sales representatives to market its new toys to retail stores. Under what circumstances should it pay a commission for every order of toys promoted by its sales representatives, and under what circumstances should it pay the sales representatives an hourly wage?
8. Analyze the labor supply schedules for Joshua and Scott below.

Wage	Hours Worked by Scott	Hours Worked by Joshua
$5	5	0
$8	10	8
$12	20	15
$15	30	25
$18	40	35
$20	45	33
$25	50	30

   a. Draw the labor supply schedules for Joshua and Scott.
   b. How does Scott's marginal benefit from more leisure compare with Joshua's?
   c. At what point does the income effect begin to outweigh the substitution effect for Joshua? Explain.

9. A competitive firm has the production function shown in the table below. Calculate the marginal product of labor and draw the marginal revenue product schedule when the market price of the good this firm produces equals $10 per ton.

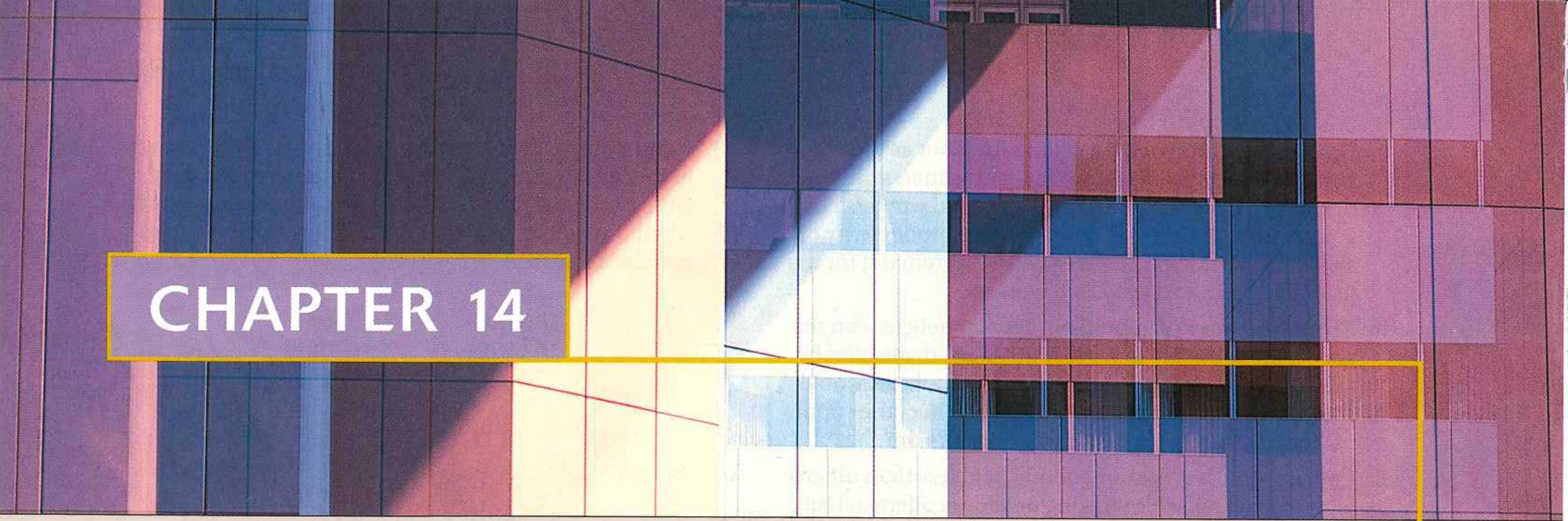
Quantity of Labor	Tons of Output
1	10
2	18
3	25
4	30
5	34
6	37
7	38

   a. If the wage is $40, how many workers will the firm hire? Explain the reasoning behind the firm's decision.
   b. If the price of the product this firm produces goes up to $15 per ton, how many workers will the firm hire? (Assume the market wage stays the same.) Why does it make sense for the firm to hire more workers?

10. Suppose a firm with some market power faces a downward-sloping demand curve for the product it produces. Given the following information on demand, complete the table below and draw the resulting demand curve for labor. If the hourly wage is $30, how many workers will this firm hire?
11. The government of Firmland wants to favor firms, and it is considering implementing a maximum wage. As an economic adviser to the government of Firmland, explain (verbally and graphically) the consequences of the maximum wage in the competitive labor market. Make sure your explanation includes the gains or losses to firms, workers, and the people of Firmland as a whole.

Problem 10

Quantity of Labor	Quantity of Output	Marginal Product of Labor	Price of Output	Total Revenue	Marginal Revenue	Marginal Revenue Product of Labor
10	100		9			
20	180		8			
30	240		7			
40	280		6			
50	300		5			
60	310		4			



CHAPTER 14

Taxes, Transfers, and Income Distribution

In 1996, President Clinton signed one of the most significant and controversial pieces of legislation of his eight years in office, promising to "end welfare as we know it." The bill allowed state governments to limit the period of time during which a poor mother with children could receive welfare payments without working. The result of the bill was that millions of poor people in the United States left the welfare rolls. The legislation was popular because it removed people from welfare and placed many in jobs. But it was severely criticized because it seemed too harsh.

There is a wide range of individual opinion about the causes of income inequality and what government should do about it. Many feel compassion and a moral obligation to help the very poor. Others feel it is unfair that some people make more in one day than others do in an entire year. Others see nothing unfair about a very unequal income distribution as long as there is equality of opportunity. Still others worry that a very unequal distribution of income can cause social unrest and deter a society from other goals, including an efficient economy.

Throughout the twentieth century all the world's democracies have chosen to set up government-run redistribution systems aimed at either reducing income inequality or helping the poor. Taxes and transfers lie at the heart of any government redistribution system. By taxing individuals who are relatively well off and making transfer payments to those who are relatively less well off, the aim is to make income distribution less unequal.

The purpose of this chapter is to provide an economic analysis of taxes and transfers. We begin the chapter with an analysis of the tax system, which is used to pay not only for transfer payments to the poor but also for government spending of all types—military, police, road building, schools. We then go on to consider transfers, such as welfare payments to the poor and social security payments to the elderly. Finally, we examine the actual distribution of income and discuss how it has been affected by the tax and transfer system in the United States.

This is an exciting time to study tax and income distribution policy. We are now observing some of the effects of the major welfare legislation enacted in the late 1990s. Passionate debates rage about whether taxes should be decreased or whether the rich should pay more or less in taxes. A major issue in the 2000 presidential election was whether tax rates should be reduced, and when he was elected, President Bush did put forward a tax cut that was passed by Congress—the Economic Growth and Tax Relief Reconciliation Act of 2001. What effects does lowering tax rates have on the economic behavior of people? How does it affect income distribution?

This chapter endeavors to provide you with some economic principles that will help you form and defend your opinions about these controversial matters.

The Tax System

We first consider the several different types of taxes used in the United States. Then we review the effects of these taxes and consider some proposals for reforming the tax system.

The major types of taxes that exist in the United States are the *personal income tax* on people's total income, the *payroll tax* on wage and salary income, the *corporate income tax* on corporate profit income, *excise/sales taxes* on goods and services purchased, *estate* and *gift taxes* on inheritances and gifts from one person to another, and *tariffs*, which are taxes on goods imported into the country. In addition, many local governments raise revenue through *property taxes*.

As shown in Figure 14.1, the personal income tax and the payroll tax are by far the largest sources of tax revenue for the federal government. Together they account for nearly 85 percent of federal tax revenue. Hence, we focus most of our attention on these two taxes in the following discussion.

The Personal Income Tax

personal income tax: a tax on all forms of income an individual or household receives.

The **personal income tax** is a tax on all the income an individual or household receives, including wage and salary income, interest and dividend income, income

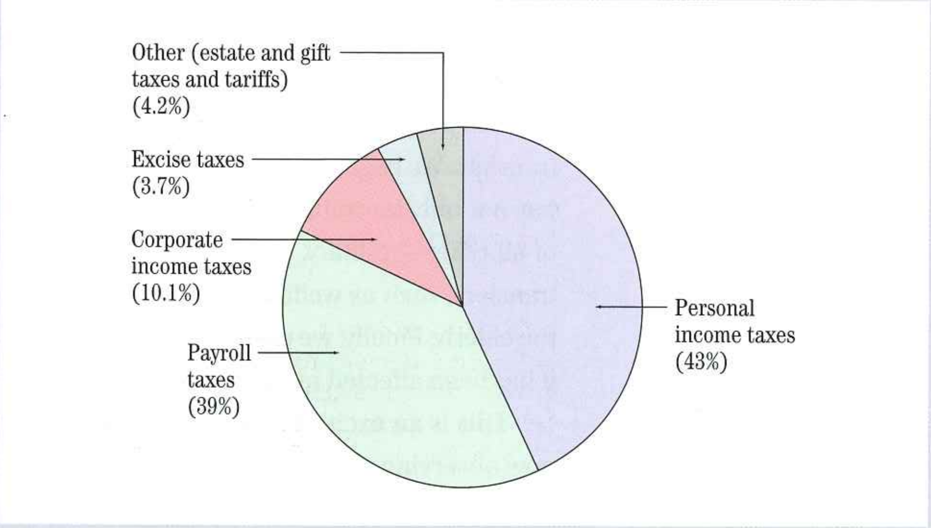


Figure 14.1
Taxes Paid to the Federal Government
Nearly 85 percent of federal taxes comes from the personal income tax and the payroll tax.

from a small business, rents on property, royalties, and capital gains. (A *capital gain* is the increase in the value of an asset like a corporate stock. When the asset is sold, the capital gain—equal to the difference between the original purchase price and the selling price of the asset—is treated as income and is taxed.) The personal income tax was introduced in 1917 in the United States, soon after the ratification of the Sixteenth Amendment to the U.S. Constitution, which authorized income taxes. Most states have now joined the federal government and have enacted a personal income tax; we focus our attention on the personal income tax collected by the federal government.

taxable income: a household's income minus exemptions and deductions.

Computing the Personal Income Tax. To explain the economic effects of the personal income tax, we must examine how people actually compute their own tax. The amount of tax a household owes depends on the tax rate and the amount of taxable income. **Taxable income** is defined as a household's income minus certain exemptions and deductions. An *exemption* is a dollar amount that can be subtracted for each person in the household. *Deductions* are other items—such as interest payments on a home mortgage, charitable contributions, and moving expenses—that can be subtracted.

Consider, for example, the Lee family, which has four members: a wife, a husband, and two children. Suppose the Lees can subtract $3,100 as a personal exemption for each of the four people in the family, for a total of $12,400, and are entitled to a deduction of $9,700. Thus, they can subtract a total of $22,100 ($12,400 + $9,700) from their income. Suppose that the husband and wife together earn a total income of $75,000. Then their taxable income is $52,900 ($75,000 − $22,100).

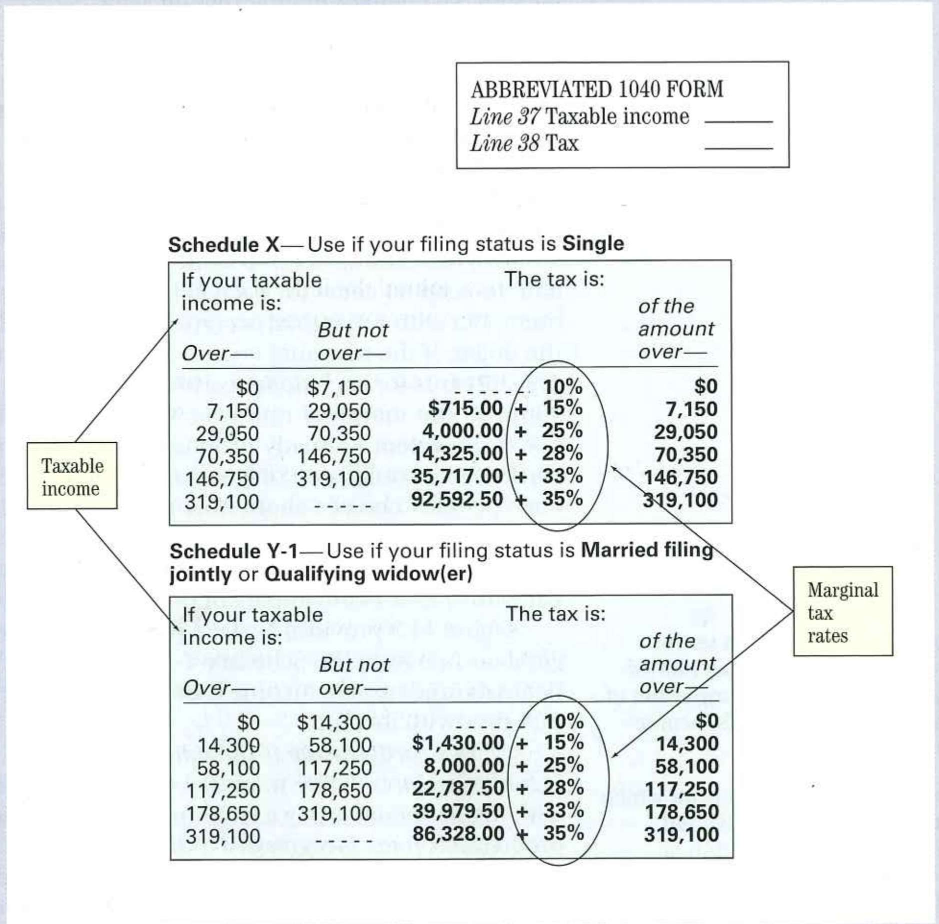
Now let us see how we combine taxable income with the tax rate to compute the tax. Figure 14.2 shows two different tax rate schedules that appeared in a recent IRS 1040 form. The tax rate schedule labeled "Schedule X" in the figure is for a taxpayer who is single; the tax rate schedule labeled "Schedule Y-1" is for two married taxpayers who are paying their taxes together. The first two columns give a range for taxable income, or the "amount on Form 1040, line 37." The next two columns tell how to compute the tax. The percentages in the tax rate schedule are the tax rates.

Look first at Schedule Y-1; the 10 percent tax rate in the schedule applies to all taxable income up to $14,300, at which point any additional income up to $58,100 is taxed at 15 percent. Any additional income over $58,100 but less than $117,250 is taxed at 25 percent, and so on for tax rates of 28 percent, 33 percent, and 35 percent. Each of the rows in these schedules corresponds to a different tax rate; the range of taxable income in each row is called a **tax bracket.**

tax bracket: a range of taxable income that is taxed at the same rate.

As an example, let us compute the Lees' tax. Recall that their taxable income is $52,900. They are married and filing jointly, so we look at Schedule Y-1. We go to the second line because $52,900 is between $14,300 and $58,100. In other words, the Lees are in the 15 percent tax bracket. We find that they must pay $1,430 plus 15 percent of the amount their income is over $14,300—that is, plus .15 × ($52,900 − $14,300) = $5,790. Thus, the amount of tax they must pay is $1,430 + $5,790 = $7,220.

Now consider what happens when the Lees' income changes. Suppose that one of the Lees decides to earn more income by working more hours and the Lees' income rises by $2,800. Thus, their taxable income rises from $52,900 to $55,700. Now what is their tax? Again looking at Schedule Y-1, we see that the tax is $1,430 plus .15 × ($55,700 − $14,300) = $6,210. Thus, the Lees' tax has increased from $7,220 to



ABBREVIATED 1040 FORM
Line 37 Taxable income ______
Line 38 Tax ______

Schedule X—Use if your filing status is **Single**

If your taxable income is: Over—	But not over—	The tax is:		of the amount over—
$0	$7,150	------	10%	$0
7,150	29,050	$715.00	+ 15%	7,150
29,050	70,350	4,000.00	+ 25%	29,050
70,350	146,750	14,325.00	+ 28%	70,350
146,750	319,100	35,717.00	+ 33%	146,750
319,100	-----	92,592.50	+ 35%	319,100

Schedule Y-1—Use if your filing status is **Married filing jointly** or **Qualifying widow(er)**

If your taxable income is: Over—	But not over—	The tax is:		of the amount over—
$0	$14,300	------	10%	$0
14,300	58,100	$1,430.00	+ 15%	14,300
58,100	117,250	8,000.00	+ 25%	58,100
117,250	178,650	22,787.50	+ 28%	117,250
178,650	319,100	39,979.50	+ 33%	178,650
319,100	-----	86,328.00	+ 35%	319,100

Figure 14.2
Two Tax Rate Schedules from the 1040 Form
The tables show how to compute the tax for each amount of taxable income. Observe how the marginal rates rise from one tax bracket to the next.

$7,640, or $420, as their income rose by $2,800. Observe that the tax rose by exactly 15 percent of the increase in income.

The Marginal Tax Rate. The amount by which taxes change when one's income changes is the **marginal tax rate.** It is defined as the change in taxes divided by the change in income. In examining how the Lees compute their tax, we have discovered that their marginal tax rate is 15 percent. In other words, when their income increased, their taxes rose by 15 percent of the increase in income. As long as they stay within the 15 percent tax bracket, their marginal tax rate is 15 percent.

marginal tax rate: the change in total tax divided by the change in income.

Observe that the marginal tax rate depends on one's income. The marginal rate varies from 10 percent for low incomes up to 35 percent for very high incomes. Suppose that one of the Lees did not work and that their taxable income was $12,900 rather than $52,900. Then they would be in the 10 percent bracket and their marginal tax rate would be 10 percent.

average tax rate: the total tax paid divided by the total taxable income.

In contrast to the marginal tax rate, the **average tax rate** is the total tax paid divided by the total taxable income. For example, the Lees' average tax rate before we considered changes in their income was $\frac{\$7,220}{\$52,900} = .136$, or 13.6 percent, lower than the 15 percent marginal tax rate. In other words, the Lees pay 13.6 percent of their total taxable income in taxes but must pay 15 percent of any additional income in taxes. The average tax rate is less than the marginal tax rate because the Lees pay only 10 percent on the first $14,300 of taxable income.

Economists feel that the marginal rate is important for assessing the effects of taxes on individual behavior. Their reasoning can be illustrated with the Lees again. Suppose that the Lees' marginal tax rate was 10 percent rather than 15 percent. Then, if one of the Lees decided to work an additional half day a week, the family would be able to keep 90 cents for each extra dollar earned, sending 10 cents to the government. But with a marginal tax rate of 15 percent, the Lees could keep only 85 cents on the dollar. If the marginal tax rate for the Lees was 35 percent, then they could keep only 65 cents for each dollar earned. To take the example to an even greater extreme, suppose the marginal rate was 91 percent, which was the highest marginal rate before President Kennedy proposed reducing tax rates. Then, for each extra dollar earned, one could keep only 9 cents! Clearly, the marginal tax rate is going to influence people's choices about how much to work if they have a choice. The marginal tax rate has a significant effect on what people gain from working additional hours. This is why economists stress the marginal tax rate rather than the average tax rate when they look at the impact of the personal income tax on people's behavior.

Figure 14.3 provides a visual perspective on marginal tax rates. It plots the marginal tax rate from IRS Schedule Y-1 in Figure 14.2; the marginal tax rate is on the vertical axis, and taxable income is on the horizontal axis. Observe how the marginal tax rate rises with income.

progressive tax: a tax for which the amount of an individual's taxes rises as a proportion of income as the person's income increases.

regressive tax: a tax for which the amount of an individual's taxes falls as a proportion of income as the person's income increases.

proportional tax: a tax for which the amount of an individual's taxes as a percentage of income is constant as the person's income rises.

*A tax is **progressive** if the amount of the tax as a percentage of income rises as income increases.* If the marginal tax rate rises with income—in which case people with higher incomes pay a larger percentage of their income in taxes—then the tax is progressive. *A tax is **regressive** if the amount of the tax as a percentage of income falls as income rises.* An income tax would be regressive if the marginal tax rate declined as income rose, or if people with high incomes could use deductions or other schemes to reduce the tax they paid to a smaller percentage of income than people with lower incomes paid. *A tax is **proportional** if the amount of the tax as a percentage of income is constant as income rises.*

Zero Tax on Low Incomes. In assessing how progressive the income tax is, one needs to remember that the taxes are based on taxable income, which is less

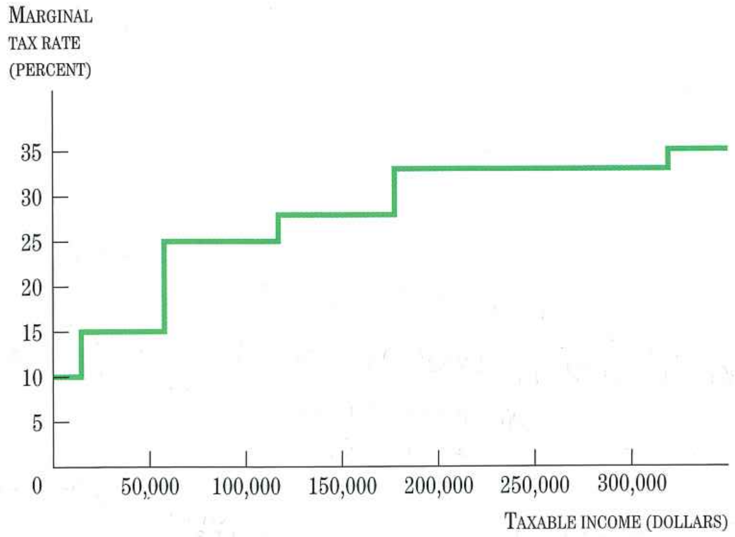


Figure 14.3
Marginal Tax Rates
As an example, the marginal tax rates from the IRS tax rate schedule Y-1 are plotted. The marginal tax rate is the change in the amount of tax paid for an extra dollar earned. The marginal tax rate increases with income. Each step takes the taxpayer to a higher tax bracket. Thus, higher-income people have a higher marginal tax rate than lower-income people. Under a flat tax, the marginal tax rate would be constant for all taxable income levels.

than the income a household actually receives. Taxable income can be zero even if a household's income is greater than zero. For example, if the Lee family earned only $22,100 for the year, then their taxable income would be zero, because $22,100 equals the sum of their exemptions and deductions. Thus, they would not have to pay any personal income tax on incomes up to $22,100, according to the tax rate schedule. In general, the personal income tax is zero for household incomes up to the sum of the exemptions and deductions.

flat tax: a tax system in which there is a constant marginal tax rate for all levels of taxable income.

A **flat tax** occurs when the marginal tax rates are constant for all levels of taxable income, in which case the line in Figure 14.3 would become flat. Even a flat rate tax system would have a degree of progressivity: The tax paid would rise as a percentage of income from zero (for workers below the sum of exemptions and deductions) to a positive amount as income rises.

The Payroll Tax

payroll tax: a tax on the wages and salaries of individuals.

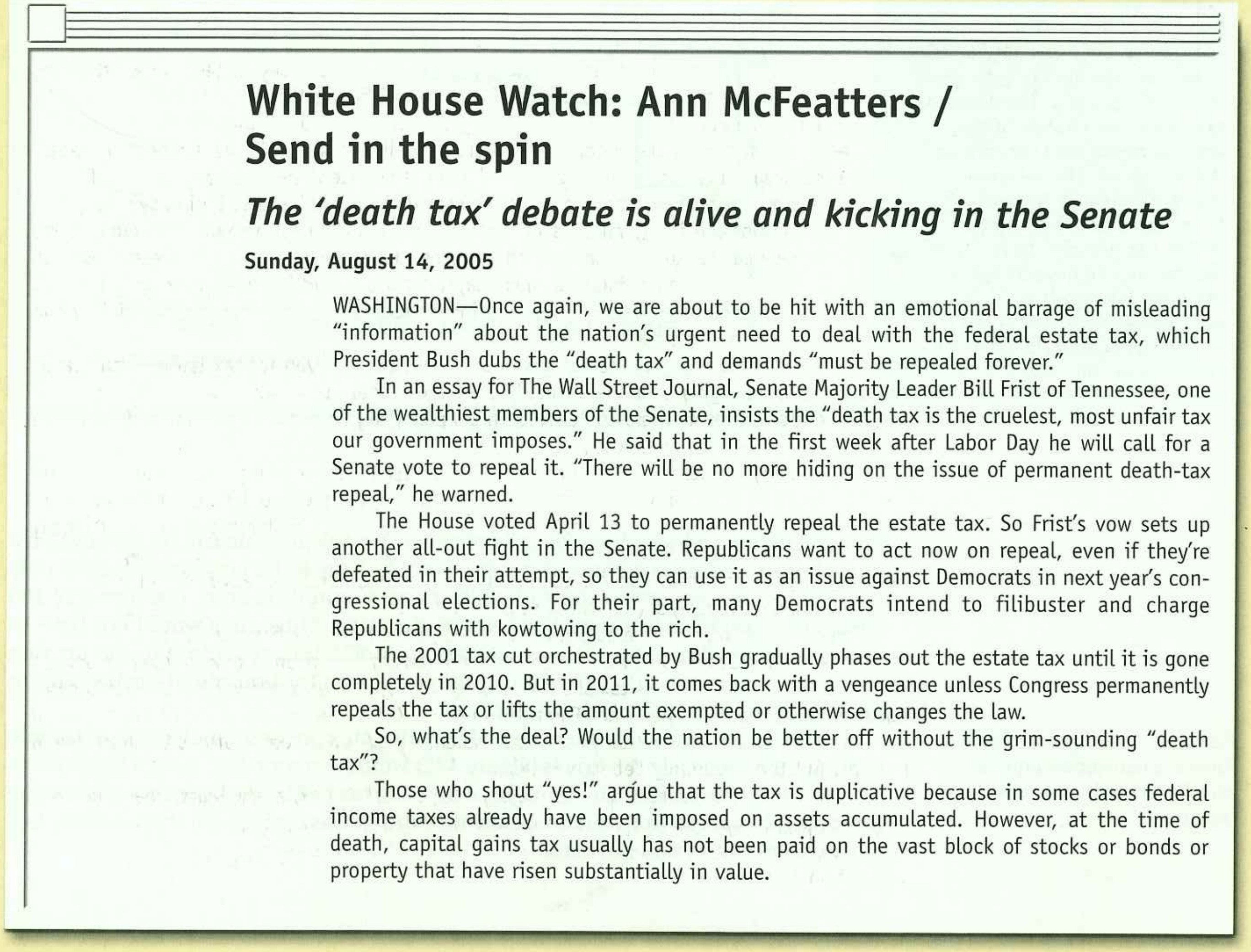
The **payroll tax** is a tax on the wages and salaries of individuals; the payroll tax goes to finance social security benefits, Medicare, and Unemployment insurance.

Payroll taxes are submitted to the government by employers. For example, the Lees' employers must submit 15.3 percent of the Lees' wage and salary income to the federal government. Thus, the payroll tax on the Lees' wage and salary income of $75,000 would be $11,475 (that is, .153 × $75,000), more than the total that the Lees would pay in personal income taxes!

The tax law says that half of the 15.3 percent payroll tax is to be paid by the worker and half is to be paid by the employer. Thus, the Lees would be notified of only half of the payroll tax, or $5,737.50, even though their employer sent $11,475 to the government. If a person is self-employed—a business consultant, say, or a freelance editor—then the person pays the full 15.3 percent, because a self-employed person is both the employee and the employer. One of the most important things to understand about the payroll tax is that, as we will soon prove, its economic effects

READING THE NEWS ABOUT... The "Death Tax" Debate

Repeal of the federal estate tax has been a topic of frequent debate since it was included as part of the Bush tax cuts enacted in 2001. Under this legislation, the estate tax is to be gradually phased out until it is completely gone in 2010, but unless further measures are taken to repeal the tax permanently, the estate tax will return in 2011 under sunset provisions of that law. As with most tax policy, the estate tax issue is not simple or clear-cut, and it has, in fact, generated a great deal of heat, as described in the article below.



White House Watch: Ann McFeatters / Send in the spin

The 'death tax' debate is alive and kicking in the Senate

Sunday, August 14, 2005

WASHINGTON—Once again, we are about to be hit with an emotional barrage of misleading "information" about the nation's urgent need to deal with the federal estate tax, which President Bush dubs the "death tax" and demands "must be repealed forever."

In an essay for The Wall Street Journal, Senate Majority Leader Bill Frist of Tennessee, one of the wealthiest members of the Senate, insists the "death tax is the cruelest, most unfair tax our government imposes." He said that in the first week after Labor Day he will call for a Senate vote to repeal it. "There will be no more hiding on the issue of permanent death-tax repeal," he warned.

The House voted April 13 to permanently repeal the estate tax. So Frist's vow sets up another all-out fight in the Senate. Republicans want to act now on repeal, even if they're defeated in their attempt, so they can use it as an issue against Democrats in next year's congressional elections. For their part, many Democrats intend to filibuster and charge Republicans with kowtowing to the rich.

The 2001 tax cut orchestrated by Bush gradually phases out the estate tax until it is gone completely in 2010. But in 2011, it comes back with a vengeance unless Congress permanently repeals the tax or lifts the amount exempted or otherwise changes the law.

So, what's the deal? Would the nation be better off without the grim-sounding "death tax"?

Those who shout "yes!" argue that the tax is duplicative because in some cases federal income taxes already have been imposed on assets accumulated. However, at the time of death, capital gains tax usually has not been paid on the vast block of stocks or bonds or property that have risen substantially in value.

do not depend on who is legally required to pay what share of the tax; only the total 15.3 percent matters.

Other Taxes

All other federal taxes together amount to less than one-fifth of total revenue. **Corporate income taxes** are taxes on the accounting profits of corporations. Currently the corporate tax rate ranges from 15 percent to 38 percent, depending on the level of earnings.

corporate income tax: a tax on the accounting profits of corporations.

Supporters of repeal say estate taxes are unfair to farm families and small businesses. But the Tax Policy Center estimates that last year estate taxes were paid on only 440 farms and small businesses out of the thousands of farms and small businesses left to heirs, and that taxes paid averaged less than 20 percent of the value of the estate. Also, there are special rules for estates with farms and businesses that reduce taxes owed.

Those who shout "no!" to repeal say the number of families actually subject to the estate tax is minuscule. The Congressional Research Service notes IRS figures that show that of the 2.4 million people who die in this country each year, only 1.3 percent of their estates owe any estate tax.

The "no" side makes the point that billions of dollars lost to the government by repeal would either mean cuts in federal programs or higher taxes elsewhere. The Congressional Budget Office estimates lost revenue from repealing the estate tax would be $380 billion over 10 years.

In a little-noticed irony, millions of American who pay no estate tax now could be hit with heavy new capital gains taxes on inheritances, depending on what Congress does. And the complications of determining how much capital gains would be due on property held for years by someone now deceased would be mind-boggling. Nobody in his or her right mind would want to be executor of a will.

The astonishing thing is that this would be so arduous, even for tax lawyers, and taxes could be so much higher that there would be a huge uproar at the same time Bush is promising, as he did this past week, to "develop a simpler [tax] code that's a fairer code and one that encourages economic growth."

Without any doubt, the coming debate will be mean and confusing. The argument of the American Conservative Union, for repeal, is: "Everything you have worked hard for your entire life, everything you wanted to leave to your children and grandchildren to keep your legacy alive, will again be taxed. And this time the tax rate will be that of a loan shark." **FactCheck.org**, which calls itself a nonpartisan, nonprofit, consumer advocate for voters and takes no stand on repeal, says nothing in that statement is true.

Supporters of keeping the estate tax argue that the true beneficiaries of repeal would be America's wealthiest families, those with many millions of dollars and flanks of lawyers able to figure out how to set up new tax shelters such as family limited partnerships and avoid capital gains taxes. Calling repeal the "Paris Hilton Relief Act," they claim the estate tax is one of the fairest, most progressive taxes, and that repeal would lower contributions to charity by 6 percent. But that's a highly debatable claim.

There is little doubt that the inordinately complex estate tax, at the least, needs an overhaul with higher exemptions. A million dollars ain't what it used to be.

But the vitriolic, misleading, partisan debate in the Senate we're about to have to endure will not do anyone any good.

excise tax: a tax paid on the value of goods at the time of purchase.

sales tax: a type of excise tax that applies to total expenditures on a broad group of goods.

Excise taxes are taxes on goods that are paid when the goods are purchased. The federal government taxes several specific items, including gasoline, tobacco, beer, wine, and hard liquor. A **sales tax** is a type of excise tax that applies to total expenditures on a broad group of goods. For example, if your expenditures on many different goods at a retail store total $100 and the sales tax rate is 5 percent, then you pay $5 in sales tax. There is no national sales tax in the United States, but sales taxes are a major source of revenue for many state and local governments.

Finally, the federal government raises revenue by imposing tariffs on goods as they enter the United States. Until the Sixteenth Amendment was ratified and the personal income tax was introduced, tariffs were the major source of revenue for the U.S. government. Now revenue from tariffs is a minor portion of total revenue.

property tax: a tax on the value of property owned.

Local governments rely heavily on **property taxes**—taxes on residential homes and business real estate—to raise revenue. Recall that income taxes—both personal and corporate—are also used at the state level.

The Effects of Taxes

The purpose of most of the taxes just described is to raise revenue, but the taxes have effects on people's behavior. To examine these effects, let us start with a tax we looked at before in Chapter 7: a tax on a good or service.

The Effect of a Tax on a Good. Recall that a tax on a good adds the amount of the tax to the marginal cost of the seller of the good. For example, a tax of $1 on a gallon of gasoline will add $1 to the marginal cost of each gallon. An increase in tax therefore shifts the supply curve up by the amount of the tax, a result shown in Figure 7.10 on page 187. Once the supply curve shifts, the ultimate impact on price and quantity will depend on the price elasticities of supply and demand.

The four panels of Figure 14.4 are designed to enable us to show how the price elasticity of demand and the price elasticity of supply determine the impact of the tax. In each of the four panels of the figure, the supply curve shifts up due to a tax of the same amount, shown by the blue arrow to the left of each vertical axis. And in each of the four panels, the equilibrium price rises and the equilibrium quantity falls. The equilibrium quantity falls because people reduce the quantity demanded of the good as its price rises because of the tax. The decline in the equilibrium quantity creates a loss of consumer surplus plus producer surplus, which we have called the deadweight loss from the tax. The size of the deadweight loss and the relative size of the impact on the price and the quantity are different in each panel of Figure 14.4 because the supply curve and the demand curve have different price elasticities.

One key point illustrated in Figure 14.4 is that *when the price elasticity of demand or the price elasticity of supply is very low, the deadweight loss from the tax is small.* This is shown in the two graphs in the left part of Figure 14.4, which have either a low elasticity of demand (top left) or a low elasticity of supply (bottom left). In either case, the deadweight loss is small compared with that in the graphs at the right, which have higher elasticities.

The intuitive reason why low elasticities result in small deadweight losses is that the quantity of the good does not change very much when the price changes. Recall that a low price elasticity of demand means that quantity demanded is not very sensitive to a change in the price, as, for example, in the case of a good like salt, which has few substitutes. A low elasticity of supply means that there is only a small change in the quantity supplied when the price changes. Thus, in the case of low elasticities, there is only a small difference between the efficient quantity of production and the quantity of production with the tax. There is little loss of efficiency. On the other hand, *when the price elasticity of demand or the price elasticity of supply is very high, the deadweight loss from the tax will be relatively large.* Here changes in price have big effects on either the quantity demanded or the quantity supplied, and the deadweight loss is large.

tax incidence: the allocation of the burden of the tax between buyer and seller.

The price elasticities of supply and demand also affect how much the price changes in response to a tax. If the price rises by a large amount, then the tax is passed on to buyers in higher prices, and the burden of the tax falls more on buyers. If the price rises little or not at all, then the seller absorbs the burden of the tax, and most of the tax is not passed on to buyers. **Tax incidence** refers to who actually bears

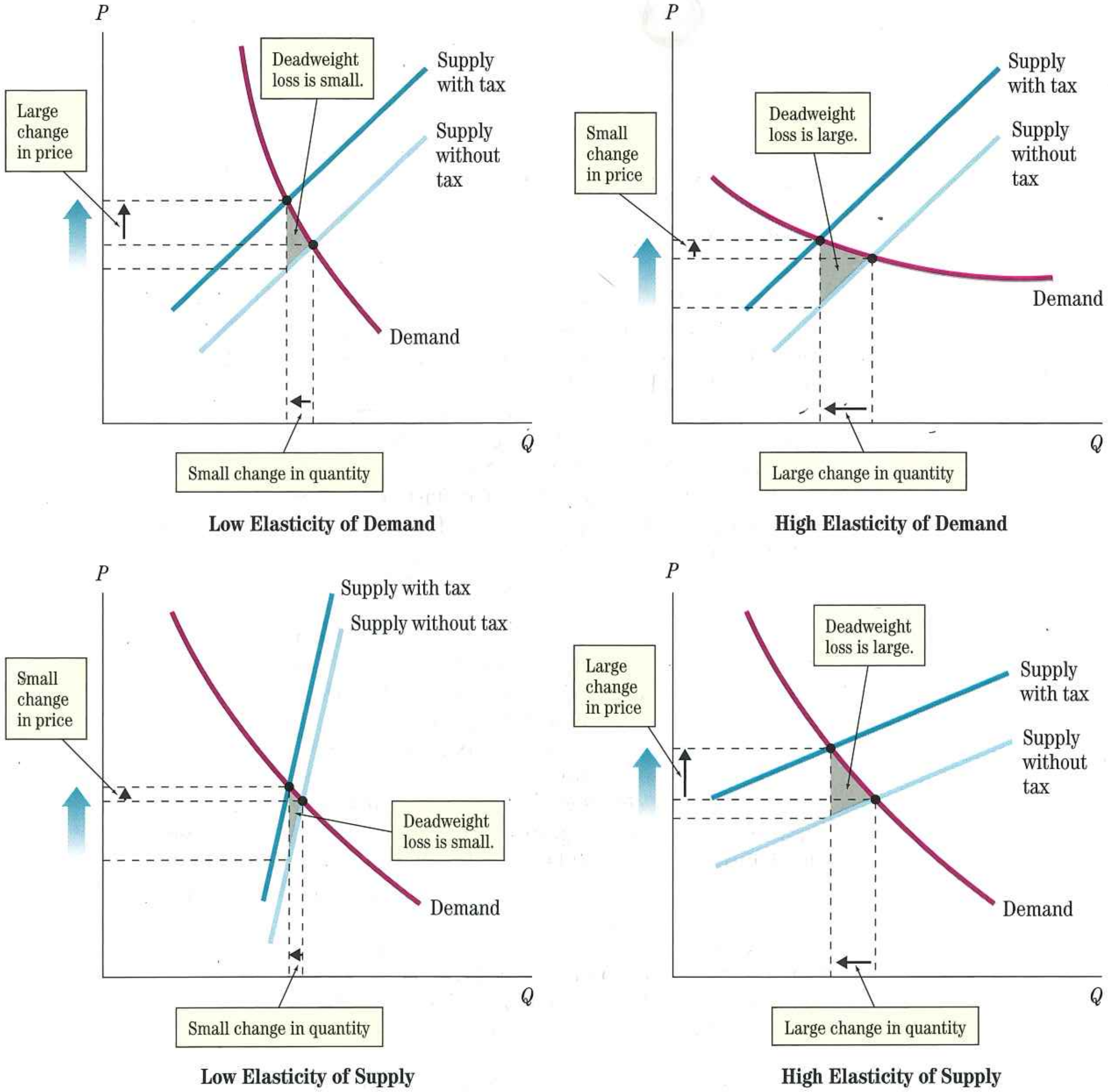


Figure 14.4
How Elasticities Determine the Effects of Taxes

1) *Deadweight loss effects:* When price elasticities are low, as in the left graphs, the deadweight loss is small and the change in equilibrium quantity is small. When price elasticities are high, as in the right graphs, the deadweight loss is large and the change in equilibrium quantity is large. 2) *Tax incidence and price effects:* When the price elasticity of demand is low or the price elasticity of supply is high, the tax is largely passed on to the consumer in higher prices. In contrast, when the price elasticity of demand is high or the price elasticity of supply is low, the burden of the tax falls on the producer because there is little price change.

the burden of the tax, the buyers or the sellers. By comparing the graphs in Figure 14.4, we see that *the smaller the price elasticity of demand and the larger the price elasticity of supply, the greater the rise in the price.* Comparing the upper two graphs of Figure 14.4, we see that the price rise is larger on the left, where the elasticity of demand is lower. The price elasticity of demand for cigarettes is very low. In 2004, the state of Michigan increased its tax on a pack of cigarettes by 75 cents, from $1.25 to $2.00. State policymakers predicted that the tax increase would raise $295 million per year in revenue and would cause people to stop smoking. They estimated that practically doubling the state's tax on cigarettes would cause cigarette sales to go down by 14 percent. Since demand for cigarettes is price inelastic, we can predict that the tax will largely be passed on to cigarette consumers, who will pay higher prices for cigarettes. The sellers can pass on the higher price to consumers when the elasticity of demand is low.

Comparing the lower two graphs of Figure 14.4, we see that the price rise is smaller on the left, where the elasticity of supply is lower. Thus, taxing a good like land, which has a low elasticity of supply, will not affect the price very much. The suppliers of the land bear the burden of the tax.

▩ **Effects of the Personal Income Tax.** We can apply our analysis of a tax on gasoline or salt to any other tax, including the personal income tax. The personal income tax is a tax on *labor* income (wages and salaries) as well as on *capital* income (interest, dividends, small business profits). However, labor income is by far the larger share of most people's income: For all 1040 forms filed, wages and salaries are over 75 percent of total income. Thus, we first focus on the personal income tax as a tax on labor income.

The analysis of the personal income tax is illustrated in Figure 14.5. Because the personal income tax is a tax on labor income, we need a model of the labor market to examine the effects of the tax. Figure 14.5 shows a labor demand curve and a labor supply curve. The wage paid to the worker is on the vertical axis, and the quantity of labor is on the horizontal axis. Figure 14.5 shows that the personal income tax shifts up the labor supply curve. The size of the upward shift depends on the marginal tax rate. For example, if a person were to supply more time working, the income received from work would be reduced by the marginal tax. If the person was in the 15 percent bracket, the income received from working would be 85 cents for each extra dollar earned working. Thus, to supply exactly the same quantity as without the tax, people require a higher wage. Because the wage paid to the worker is on the vertical axis, the labor supply curve shifts up to show this.

As the labor supply curve shifts up, the equilibrium quantity of labor declines. Thus, we predict that an income tax will reduce the amount of work. The reduced amount of work will cause a deadweight loss just like that caused by the tax on a commodity. The size of the decline in hours of work will depend on the labor supply and labor demand elasticities. The higher the labor supply elasticity, the greater the reduction in the quantity of labor supplied in response to the personal income tax.

Economists disagree about the size of the labor supply elasticity. One thing that is sure is that the elasticity is different for different people. For example, the labor supply elasticity appears to be quite high for second earners in a two-person family such as the Lees. If elasticity is high, a high marginal tax rate can reduce hours of work and thereby income. But if the labor supply curve has a low elasticity, there is little effect on hours of work.

▩ **The Effect of a Payroll Tax.** We can use the same type of labor market diagram to analyze a payroll tax, as shown in Figure 14.6. Clearly, the payroll tax is a tax on labor in that it applies to wages and salaries. However, in the case of the payroll

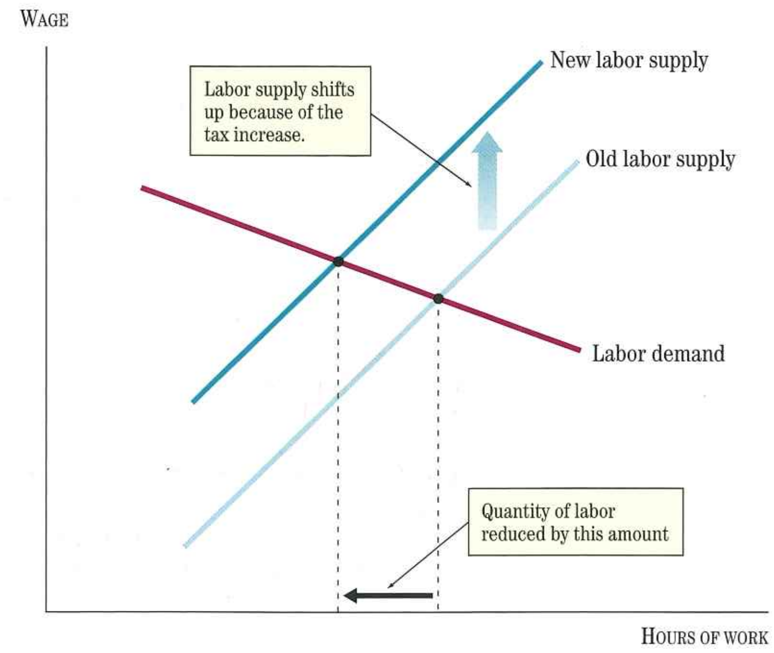


Figure 14.5
Effects of a Higher Income Tax on Labor Supply
An income tax shifts the labor supply curve up by the amount of the tax on each extra hour of work because the worker must pay part of wage income to the government and thus receives less for each hour of work. Thus, the quantity of labor supplied declines. The decline in hours worked would be small if the supply curve had a low elasticity.

tax, we need to consider that the tax is paid by both the employer and the employee, as required by law. Figure 14.6 handles the two cases.

Suppose that the wage before the tax is $10 per hour and that the payroll tax is 10 percent, or $1 per hour. The case where the tax is paid by the employee is shown on the right of Figure 14.6. This picture looks much like Figure 14.5. The labor supply curve shifts up by the amount of the tax ($1) because the worker now has to pay a tax to the government for each hour worked. In other words, the worker will supply the same amount of work when the wage is $11 and the tax is $1 as when the wage is $10 and the tax is zero.

When the labor supply curve shifts up, we see in the right-hand panel of Figure 14.6 that the equilibrium quantity of labor employed declines. Observe that the wage paid by the employer rises because the reduced supply requires a reduction in the quantity of labor demanded, which is brought about by a higher wage. However, the "after-tax wage"—the wage less the tax—declines because the tax increases by more than the wage increases.

The case where the tax is paid by the employer is shown in the left graph of Figure 14.6. In this case, the labor demand curve shifts down by the amount of the tax ($1) because the firm has to pay an additional $1 for each hour of work. When the labor demand curve shifts down, the equilibrium quantity of labor employed declines and the wage falls. Observe that the impact of the payroll tax is the same in both cases: There is a new equilibrium in the labor market with a lower wage and a lower quantity of labor.

Thus, a payroll tax has both an employment-reduction effect and a wage-reduction effect. As with any tax, the size of the quantity change and the price (wage)

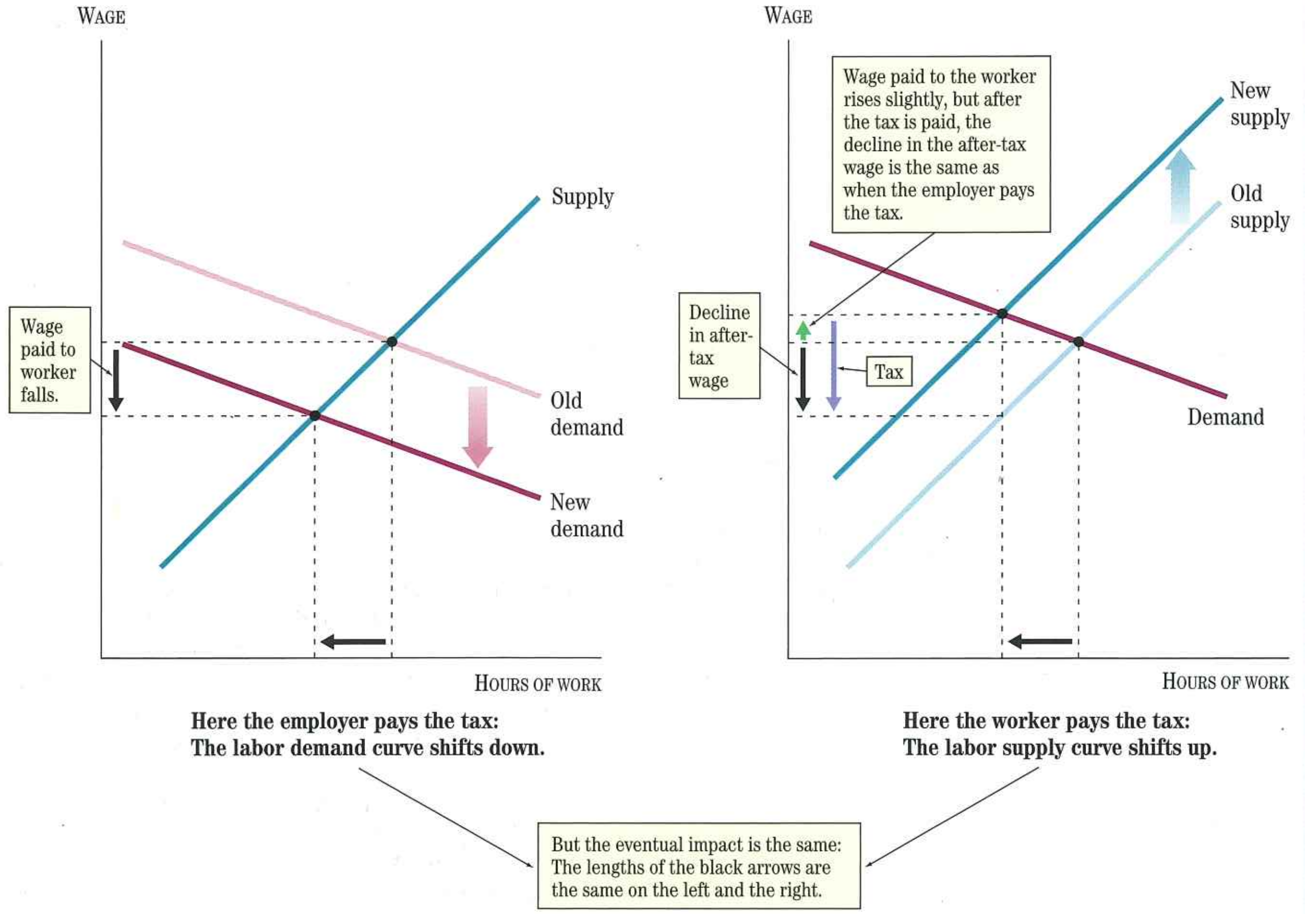


Figure 14.6
Effect of Payroll Tax

If a payroll tax is paid by the employer, the labor demand curve shifts down by the amount of the tax because the firm's labor costs increase by the amount of the tax. Thus, the quantity of labor employed declines, as does the wage paid to the worker, as shown on the left. A payroll tax paid by the employee, shown on the right, causes the labor supply curve to rise by the amount of the tax, but the effects on after-tax wages received by the worker and the quantity of work are the same as when the employer pays.

change depends on the supply and demand elasticities. For example, if the labor supply elasticity is low, there will be a small reduction in employment, but the wage will fall by a large amount. However, if the labor supply elasticity is high, there will be a large employment effect, but the wage effect will be small.

tax revenue: the tax rate times the amount subject to tax.

■ **The Possibility of a Perverse Effect on Tax Revenue.** Tax revenue received by the government is equal to the tax rate times the amount that is subject to the tax. For example, in the case of a gasoline tax, the tax revenue is the tax per gallon times the number of gallons sold. As the tax rate increases, the amount subject to the tax will fall because the higher price due to the tax reduces the quantity

Table 14.1
Tax Rates and Tax Revenue: An Example

Tax Revenues	Tax Rate	Wage	Hours Worked
$10,000	.50	$10/hour	2,000
$11,250	.75	$10/hour	1,500
$ 4,500	.90	$10/hour	500

demanded. If the quantity demanded falls sharply enough, then tax revenue could actually fall when the tax rate is increased.

The same possibility arises in the case of taxes on labor, either the payroll tax or the personal income tax. In the case of the payroll tax or the personal income tax for a worker, tax revenue is equal to the tax rate times the wage and salary income. As the tax rate rises, the amount of income subject to tax may fall if labor supply declines. Thus, in principle, it is possible that a higher tax rate could result in reduced tax revenue. For example, consider the high marginal tax rates shown in Table 14.1: 50 percent, 75 percent, and 90 percent. If labor supply declines with a higher tax rate, as assumed in the table, then tax revenue first increases as the tax rate goes from 50 to 75 percent but then declines as the tax rate goes from 75 to 90 percent.

The general relationship between tax rates and tax revenue is illustrated in Figure 14.7. As in the example of Table 14.1, tax revenue first rises and then falls as the tax rate increases. Figure 14.7 can apply to any tax on anything. At the two extremes of zero percent tax rate and 100 percent tax rate, tax revenue is zero. What happens between these two extremes depends on the elasticities. This relationship between the tax rate and tax revenue, now frequently called the Laffer curve after the economist Arthur Laffer, who made it popular in the 1980s, has long been known to economists. It implies that if the tax rate is so high that we are on the downward-sloping part of the curve, then reducing the tax rate may increase tax revenue. However, there is great debate among economists about the tax rate at which the curve bends around (40 percent? 50 percent? 90 percent?) and how it applies in different situations.

Other factors influencing tax revenue when taxes get very high are tax avoidance and tax evasion. *Tax avoidance* means finding legal ways to reduce taxes, such as buying a home rather than renting in order to have a deduction for interest payments on a mortgage. *Tax evasion* is an illegal means of reducing one's tax. For example, at high tax rates, people have incentives to evade the tax by not reporting income. Workers are tempted not to report tips. Or people resort to barter, which is difficult for the government to track down. For example, an employer may "pay" a little extra to a truck driver by allowing free use of the truck on weekends for fishing trips.

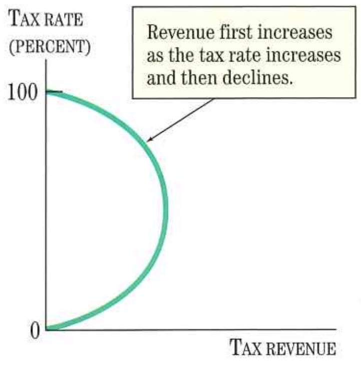


Figure 14.7
The Tax Rate and Tax Revenue
As the tax rate increases from low levels, tax revenue rises. At some point, however, the high tax rate reduces the quantity of the item that is taxed and encourages so much tax avoidance that the amount of tax revenue declines. This curve is frequently called the *Laffer curve.* The particular tax rate at which the curve bends depends on the price elasticity of the item being taxed and is a subject of great debate among economists.

Tax Policy and the Tradeoff Between Efficiency and Equality

We have observed in our analysis of each tax that the equilibrium quantity of the item taxed declines when the tax rate rises. This is where the inefficiency of the tax comes from. If the tax rate is very high, or the elasticities are very high, the inefficiency can be so severe that it could thwart one of the purposes of raising the taxes: to provide income support in order to raise the well-being of the

"And do you promise to love, honor, and cherish each other, and pay the United States government more taxes as a married couple than you would have paid if you had just continued living together?"

least well-off in the society. Why? Because the reduction in the quantity of labor supplied or goods produced could be so great that there would be less total income in the society. Thus, there would be less going to the poor even if they received a larger share of total income. In other words, there is a *tradeoff between equality and efficiency*. If one raises taxes too high for the purpose of making the income distribution more equal, the total amount of income may decline. In that event, there will be less available to redistribute.

Given these considerations, how should the tax system—the combination of all the taxes in society—be designed or improved?

First, in order to reduce deadweight loss to a minimum, the ideal tax system should tax items with small price elasticities of supply and demand rather than items with large elasticities. We know that the deadweight loss is small when elasticities are small. The optimal tax system would have tax rates inversely related to the elasticities.

Second, the ideal tax system would try to keep the marginal tax rates low and the amount that is subject to tax high. For example, we saw that deductions reduce the amount subject to personal income tax by lowering taxable income. Some deductions are put in the tax system to encourage certain activities: A deduction for research expenses may encourage firms to fund research, for example. However, the more deductions there are, the higher the tax rate has to be in order to get the same tax revenue. If people were not allowed to exclude so many items from income, a lower marginal tax rate could generate the same amount of revenue. And a lower marginal tax rate has the advantage of reducing the inefficiency of the tax.

Most tax reform efforts have involved trying to reduce the number of deductions while lowering marginal tax rates. This was the idea behind the tax reform efforts in the 1960s under President Kennedy and in the 1980s under President Reagan. In the early 2000s the marginal tax rates on all taxpayers were reduced substantially.

ability-to-pay principle: the view that those with greater income should pay more in taxes than those with less income.

Third, the ideal tax system should be as simple and as fair as possible. If a tax system is not simple, then valuable resources—people's time, computers, and so on—must be devoted to paying and processing taxes. The tax system is seen as unfair if it is regressive. Another view of fairness frequently used is the **ability-to-pay principle;** this view is that those with greater income should pay more in taxes than those with less income. The tax system is also viewed as unfair if people with the same incomes are taxed at different rates. For example, in the U.S. tax system, a married couple making more than $120,000 a year pays a higher tax than an unmarried couple with exactly the same income. This is viewed by some as unfair.

REVIEW

- Taxes are used to make transfer payments to individuals as well as to build roads and provide for education and national defense.
- Taxes cause inefficiencies in the form of reduced economic activity and deadweight loss.
- There is a tradeoff between efficiency and equality; raising taxes to reduce inequality may increase economic inefficiency and thereby reduce the amount of total income.
- To minimize the inefficiencies, items with low elasticities should be taxed more than items with high elasticities.
- In the case of taxes on labor income—a payroll tax or the income tax for most people—the amount of work declines as the tax is increased.

Transfer Payments

transfer payment: a grant of funds from the government to an individual.

means-tested transfer: a transfer payment that depends on the income of the recipient.

social insurance transfer: a transfer payment, such as social security, that does not depend on the income of the recipient.

A **transfer payment** is a payment from the government to an individual that is not in exchange for a good or service. Transfer payments can be either in cash or in kind. In-kind payments include vouchers to buy food or housing.

There are two types of government transfer payments in the United States: **means-tested transfers,** which depend on the income (the means) of the recipient and focus on helping poor people, and **social insurance transfers,** which do not depend on the income of the recipient. We will discuss each type of transfer, starting with the means-tested transfer programs.

Means-Tested Transfer Programs

Means-tested transfer payments are made to millions of people in the United States each year. The major programs are listed in Table 14.2.

family support programs: transfer programs through which the federal government makes grants to states to give cash to certain low-income families.

Medicaid: a health insurance program designed primarily for families with low incomes.

supplemental security income (SSI): a means-tested transfer program designed primarily to help the poor who are disabled or blind.

The 1996 federal welfare law (called the Personal Responsibility and Work Opportunity Reconciliation Act, or PRWORA) replaced Aid to Families with Dependent Children (AFDC)—a transfer program providing cash payments to poor families with children. Usually, AFDC has simply been called "welfare." Under the new **family support programs,** the federal government provides grants to states, which then decide which poor families are eligible. In contrast, under AFDC the federal government stipulated eligibility requirements.

Medicaid is a health insurance program that is designed primarily to pay for health care for people with low incomes. Under the new welfare law, Medicaid eligibility is based substantially on the rules for eligibility from the former AFDC program, although it is no longer linked automatically to AFDC. Once income increases to a certain level, Medicaid support stops, so that the family must find another means of obtaining health insurance. **Supplemental security income (SSI)** is a program designed to help the neediest elderly as well as poor people who are disabled or blind. About 6.6 million people receive SSI assistance, including 4.6 million disabled and 2 million aged people.

Table 14.2
Means-Tested Transfer Programs in the United States (Each of these federal programs requires that the recipient's income or assets be below a certain amount in order to receive payment.)

Family Support Programs (Welfare)	Payments to poor families with children as determined by each state
Medicaid	Health insurance primarily for welfare recipients
SSI (Supplemental Security Income)	Payments to poor people who are old, disabled, or blind
Food Stamp Program	Coupons for low-income people to buy food
Head Start	Preschool education for low-income children
Housing Assistance	Rental subsidies and aid for construction

food stamp program: a government program that provides people with low incomes with coupons (food stamps) that they can use to buy food.

Head Start: a government transfer program that provides day care and nursery school training for poor children.

housing assistance programs: government programs that provide subsidies either to low-income families to rent housing or to contractors to build low-income housing.

The **food stamp program** is a major means-tested transfer program; it makes payments to about 17 million people each year. Like Medicaid, food stamps are an in-kind payment. People are not supposed to use the coupons to buy anything but food. This is a popular program because the intent of the money is to provide nutrition and because the program is fairly inexpensive to run. The National School Lunch Program is similar to food stamps in that it aims to provide food to lower-income children. It provides school lunches for about 27 million children.

Head Start, another in-kind program, provides for preschool assistance to poor children to help them get a good start in school. It also is a popular program because there is evidence that it improves the performance, at least temporarily, of preschool children as they enter elementary school.

Housing assistance programs provide rental subsidies to people who cannot afford to buy a home. The programs sometimes provide aid to business firms that construct low-income housing. Many complain about waste and poor incentives in the housing programs and argue that these programs are in need of reform.

The Earned Income Tax Credit (EITC)

earned income tax credit (EITC): a part of the personal income tax through which people with low income who work receive a payment from the government or a rebate on their taxes.

Another program aimed at helping the poor in the United States is the **earned income tax credit (EITC).** It is like a means-tested transfer payment in that people receive a payment from the government if their income is below a certain amount. However, it is actually part of the personal income tax (the form to obtain the payment is sent to people by the IRS along with the 1040 form).

The program provides assistance to about 18 million families. The EITC is for working people whose income is below a certain level, either because their wage is very low or because they work part time. They get a refundable credit that raises their take-home pay. For example, consider the four-person Lee family again. We know that if they earn less than $22,100, they pay no income tax. However, if the Lees earn between $0 and $10,750 in wages and salary and have no other income, then the EITC will pay 40 cents for each dollar they earn up to a maximum of $4,300 per year. To make sure that the EITC does not make payments to high-income people, the payments decline if the Lees make more than $15,050. For each dollar they earn above $15,050, they lose 22 cents of their $4,300 until the benefits run out (when their income reaches $35,458 with more than one qualifying child).

Observe that the EITC raises the incentive to work for incomes up to $10,750 and reduces the incentive to work for incomes greater than $15,050 and less than $35,458. With the EITC, the marginal tax rate is effectively *negative* 40 percent for income below $10,750; that is, you *get* 40 cents rather than *pay* 40 cents for each dollar you earn. But the EITC adds 22 percent to the marginal tax rate for incomes over $15,050 up to $35,458.

Incentive Effects

The previous sections describe a variety of government programs that aim to transfer funds to the poor. As we will see, evidence suggests that these programs do have an impact in reducing income inequality. However, some people feel that the programs may create a disincentive to work, since welfare payments are reduced when income from work rises.

The top panel of Figure 14.8 illustrates the first disincentive problem. The total income of an individual is plotted against the number of hours worked. Total income consists of wage income from work plus a welfare payment. The more steeply sloped

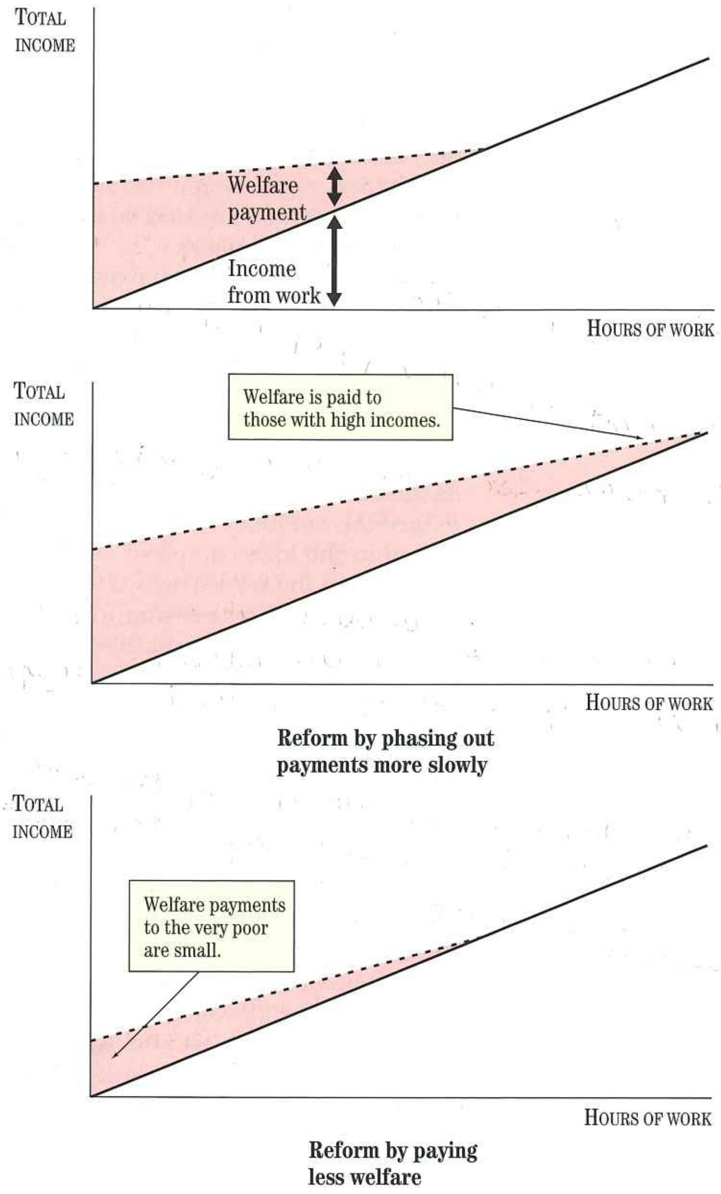


Figure 14.8
Welfare Reform to Improve Work Incentives
The top graph shows how welfare reduces the marginal earnings from working more hours because welfare payments are phased out. There are two basic approaches to reform: phasing out the payments more slowly (as in the middle graph) or lowering the welfare payment (as in the lower graph). Both have advantages and disadvantages.

solid black line shows the individual's wage income from work: The more hours worked, the more wage income the individual receives. This line intercepts the horizontal axis at zero income, so if there is no work and there is no welfare or charity, the person is in a state of extreme poverty.

The individual's total income is shown by the less steeply sloped dashed line in the top graph of Figure 14.8. It intercepts the vertical axis at an amount equal to the welfare payment the individual gets when he or she is not working at all. As the individual begins to work, the need for welfare declines, and so the welfare payment declines. Observe that the amount of the welfare payment, which is represented by the shaded gap between the steep line and the less steep line, diminishes as the hours of work increase, and finally, after a certain number of hours worked, the welfare payment disappears.

Because the welfare payment is reduced when the individual's income from work rises, it creates a disincentive. The flatter the dashed line, the greater the disincentive. For example, if someone decides to work 10 hours a week for a total of $50 per week, but the welfare payment is reduced by $30 per week, then effectively the marginal tax rate is 60 percent, high enough to discourage work.

Welfare reform endeavors to change the welfare system in order to reduce this disincentive. Looking at Figure 14.8, we see that there are two ways to make the dashed line steeper and thereby provide more incentive to work. One way is to reduce the amount of welfare paid at the zero income amount. Graphically, this is shown in the lower graph of Figure 14.8. This twists the dashed line because the intercept on the vertical axis is lower but the intersection of the dashed line and the solid line is at the same number of hours of work as in the top graph. This will increase the slope of the dashed line and therefore provide more incentive to work. But the problem with this approach is that poor people get less welfare: The poverty rate could rise.

A second way to make the dashed line steeper is to raise the place at which it intersects the black solid line, as in the middle graph of Figure 14.8. But that might mean making welfare payments to people who do not need them at all, people who earn $50,000 or $60,000 annually.

The welfare reform act signed into law by President Clinton in 1996 leaves the decision as to which welfare reform approach to take up to the states. Thus, the states have gone off in different directions, some cutting welfare checks and others raising the amount that can be earned before welfare is reduced. Some states have taken other approaches to get around the disincentive difficulty. Florida, Tennessee, and Texas require adult welfare recipients to go to work immediately. Twenty-four states require that people work after two years on welfare. Other states require that a single parent finish high school in order to get the full welfare payment. These proposals are aimed at increasing the incentive to get off welfare and go to work. Have they worked? Supporters say emphatically, yes—welfare rolls have been dramatically reduced. Critics argue that many of those previous recipients of welfare are still struggling—they may be working, but they are not earning enough to raise themselves out of poverty. These critics contend that many of these people will have trouble keeping jobs unless they have adequate support services, such as health insurance and child care.

social security: the system through which individuals make payments to the government when they work and receive payments from the government when they retire or become disabled.

Social Insurance Programs

Many transfer payments in the United States are not means-tested. The largest of these are social security, Medicare, and unemployment insurance. **Social security** is

Medicare: a government health insurance program for the elderly.

unemployment insurance: a program that makes payments to people who lose their jobs.

the system through which payments from the government are made to individuals when they retire or become disabled. **Medicare** is a health insurance program for older people. **Unemployment insurance** pays money to individuals who are laid off from work.

Social security, Medicare, and unemployment insurance are called *social insurance* because they make payments to anyone—rich or poor—under certain specific circumstances. Social security provides benefits when a worker becomes disabled or retires. Medicare provides payments when an older person requires medical care, and unemployment compensation is paid to workers when they are laid off from a job.

But these programs have features that make them much more than insurance programs. The programs have effects on income distribution because they transfer income between different groups. Consider social security and Medicare. Payroll taxes from workers pay for these programs. But the payroll taxes paid by an individual are only loosely related to the funds paid out to the same individual. In reality, each year the funds paid in by the workers are paid out to the current older people. In other words, social security is more like a transfer program from young people to older people than an insurance program.

However, because the social insurance programs are not means-tested, they also transfer income to middle-income and even wealthy individuals. In other words, they are not well targeted at the lower-income groups. For this reason, many people have suggested that these programs be means-tested. In fact, recent legislation has effectively reduced social security benefits to higher-income older people by requiring that a major part of the benefits be included in taxable income; social security benefits were formerly excluded from taxable income.

Mandated Benefits

mandated benefits: benefits that a firm is required by law to provide to its employees.

Mandated benefits occur when a firm is required by the government to provide a benefit for its workers. For example, a federal law requires firms to give unpaid leave to employees to care for a newborn baby or a sick relative. Such benefits are a cost to the firm (for example, the cost of finding and training a replacement or providing health insurance to the worker on leave). But, of course, they are a benefit to the worker. Another example of a mandated benefit is a proposal that would require firms to pay a portion of the health insurance costs of their workers.

The effects of mandated benefits can be analyzed using the supply and demand for labor diagram, much as we analyzed the effects of a payroll tax. As shown in Figure 14.9, the labor demand curve shifts down, as it did in Figure 14.6 for the employer-paid payroll tax, because the mandated benefits are a cost to the firm. But the mandated benefits provide a benefit to the workers, which shifts the labor supply curve down. The labor supply curve will probably not shift down as much as the labor demand curve because the worker probably will not value the benefit quite as much as its cost to the firm.

In any case, the new equilibrium in Figure 14.9 shows that the wage paid to the worker will fall by nearly the amount of the mandated benefit. In other words, despite the fact that the employer is "paying" for the mandated benefit, it is the worker who mainly pays. There will also be a reduction in employment.

If the workers value the benefit exactly as much as it costs the firm, then the wage will fall by the full amount of the benefit. In this case, employment will not fall at all.

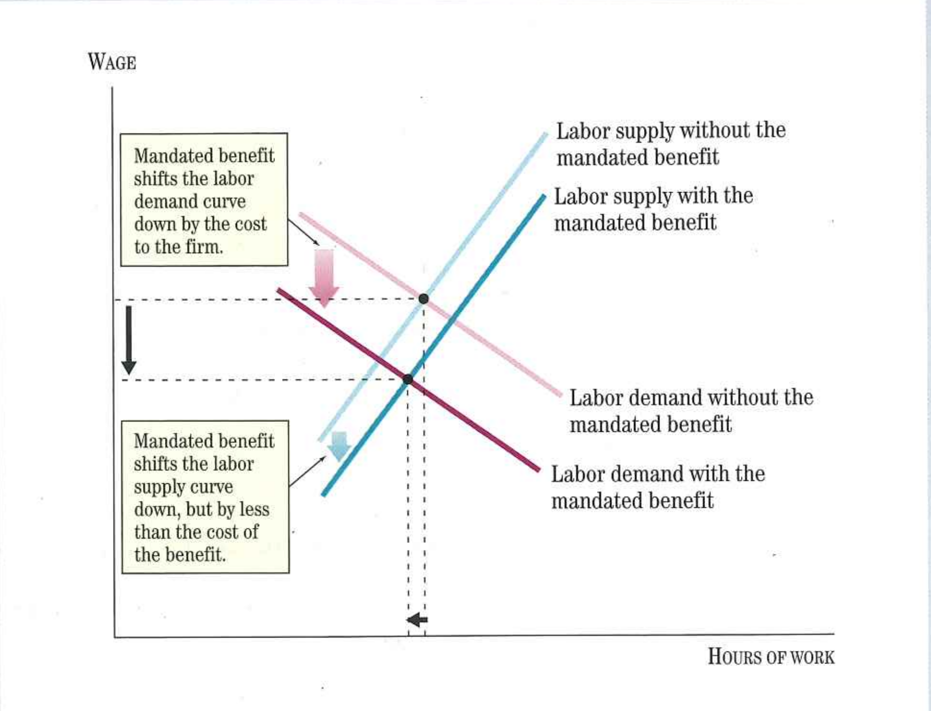

WAGE
Mandated benefit shifts the labor demand curve down by the cost to the firm.
Mandated benefit shifts the labor supply curve down, but by less than the cost of the benefit.
Labor supply without the mandated benefit
Labor supply with the mandated benefit
Labor demand without the mandated benefit
Labor demand with the mandated benefit
HOURS OF WORK

Figure 14.9
Effect of a Mandated Benefit
A mandated benefit is a cost to the firm; it shifts down the demand curve for labor just as a payroll tax does. But in the case of a mandated benefit, the labor supply curve shifts down too. Hence, the wage paid to the worker falls, as does employment.

REVIEW

- There are two major types of transfer payments. Means-tested payments—family support, Medicaid, food stamps—depend on the income of the recipient. Social insurance—social security, unemployment compensation, and Medicare—does not depend on the income of the recipient.
- Means-tested transfer payments create disincentives. The purpose of welfare reform is to reduce those disincentives.
- Under welfare reform, states are trying different methods to reduce disincentives, including work requirements and time limits on payment of benefits to welfare recipients.

The Distribution of Income in the United States

What does the distribution of income in the United States actually look like? What effect does the tax and transfer system described in the previous two sections have on income distribution? Does the United States have a less equal income distribution than other countries? What has been happening to income distribution over time? To answer these questions, we need a quantitative measure of income distribution.

The Personal Distribution of Income

Current Population Survey: a monthly survey of a sample of U.S. households done by the U.S. Census Bureau; it measures employment, unemployment, the labor force, and other characteristics of the U.S. population.

Data about people's income in the United States are collected by the Census Bureau in a monthly survey of about 70,000 households called the **Current Population Survey.** Using the information on the income of households in this survey, an estimate of the distribution of income for the entire country is made.

Economists and statisticians usually study the income distribution of families or households rather than individuals. A *family* is defined by the Census Bureau as a group of two or more people related by birth, marriage, or adoption who live in the same housing unit. A *household* consists of all related family members and unrelated individuals who live in the same housing unit. Because the members of a family or a household typically share their income, it is usually more sensible to consider families or households rather than individuals. One would not say that a young child who earns nothing is poor if the child's mother or father earns $100,000 a year. In a family without children in which one spouse works and the other remains at home, one would not say that the working spouse is rich and the nonworking spouse is poor.

quintiles: divisions or groupings of one-fifth of a population ordered by income, wealth, or some other statistic.

Because there are so many people in the population, it is necessary to have a simple way to summarize the income data. One way to do this is to arrange the population into a small number of groups ranging from the poorest to the richest. Most typically, the population is divided into fifths, called **quintiles,** with the same percentage of families or households in each quintile. For example, in Table 14.3, the 76 million families in the United States are divided into five quintiles, with 15.2 million families in each quintile. The first row shows the poorest 20 percent—the bottom quintile. The next several rows show the higher-income quintiles, with the last row showing the 20 percent with the highest incomes.

The second and third columns of Table 14.3 show how much income is earned by families in each of the five groups. The bottom 20 percent of families have incomes below $24,000, the families in the next quintile have incomes greater than $24,000 but less than $41,440, and so on. Note that the lower limit for families in the top 20 percent is $94,469. The lower limit for the top 5 percent (not shown in the table) is $164,323.

Inequality can be better measured by considering the total income in each quintile as a percentage of the total income in the country. Table 14.4 provides this information. The second column in Table 14.4 shows the income received by families in each quintile as a percentage of total income in the United States.

Table 14.3
Range of Annual Family Incomes for Five Quintiles

Quintile	Income Greater Than	Income Less Than
Bottom 20 percent	0	$24,000
Second 20 percent	$24,000	$41,440
Third 20 percent	$41,440	$63,000
Fourth 20 percent	$63,000	$94,469
Top 20 percent	$94,469	—

Source: Statistical Abstract of the United States, 2004, Table 672.

Table 14.4
Distribution of Family Income by Quintile

Quintile	Percentage of Income	Cumulative Percentage of Income
Bottom 20 percent	4.2	4.2
Second 20 percent	9.7	13.9
Third 20 percent	15.5	29.4
Fourth 20 percent	23.0	52.4
Top 20 percent	47.6	100.0

Source: Statistical Abstract of the United States, 2004, Table 670.

A quick look at Table 14.4 shows that the distribution of income is far from equal. Those in the lower 20 percent earn only 4.2 percent of total income. On the other hand, those in the top 20 percent earn 47.6 percent of total income. Thus, the amount of income earned by the rich is a large multiple of the amount of income earned by the poor.

These percentages are summed up in the third column of Table 14.4. This cumulative percentage shows that the bottom 20 percent earn 4.2 percent of the income, the bottom 40 percent earn 13.9 percent of the income, the bottom 60 percent earn 29.4 percent of the income, and the bottom 80 percent earn 52.4 percent of the income. The top 5 percent, not shown in Table 14.4, earn 20.8 percent of the aggregate income.

The Lorenz Curve and Gini Coefficient

The data in Table 14.4 can be presented in a useful graphical form. Figure 14.10 shows the cumulative percentage of income from the third column of Table 14.4 on the vertical axis and the percentage representing each quintile from the first column on the horizontal axis. The five dots in the figure are the five pairs of observations from the table. For example, point *A* at the lower left corresponds to the 4.2 percent of income earned by the lowest 20 percent of people. Point *B* corresponds to the 13.9 percent of income earned by the lowest 40 percent of people. The other points are plotted the same way. The uppermost point is where 100 percent of the income is earned by 100 percent of the people.

If we connect these five points, we get a curve that is bowed out. This curve is called the **Lorenz curve.** To measure how bowed out the curve is, we draw the solid black 45-degree line. The 45-degree line is a line of perfect equality. On that line, the lowest 20 percent earn exactly 20 percent of the income, the lowest 40 percent earn exactly 40 percent of the income, and so on. Every household earns exactly the same amount.

Lorenz curve: a curve showing the relation between the cumulative percentage of the population and the proportion of total income earned by each cumulative percentage. It measures income inequality.

The degree to which the Lorenz curve is bowed out from the 45-degree line provides a visual gauge of the inequality of income. The more bowed out the line is, the more unequal is the income distribution. The box on page 386 shows how the Lorenz curve in the United States compares with that in some other countries and with the world as a whole.

The most unequal distribution possible would occur when only one person earns all the income. In that case, the curve could be so bowed out from the

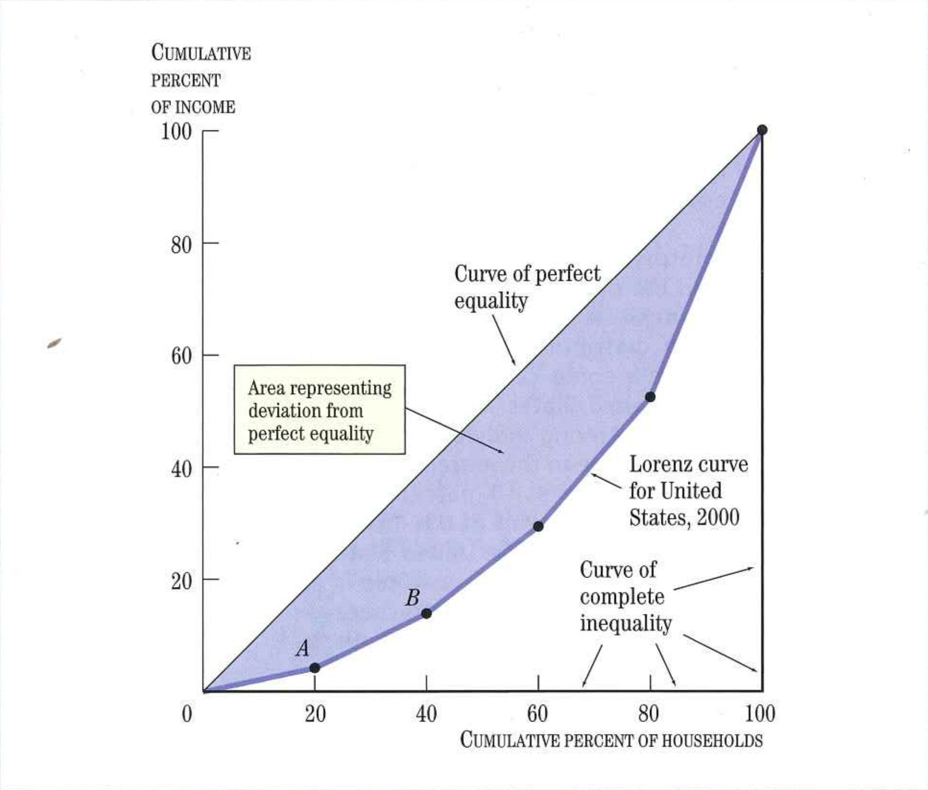


Figure 14.10
The Lorenz Curve for the United States
Each point on the Lorenz curve gives the percentage of income received by a percentage of households. The plotted points are for the United States. Point *A* shows that 4.2 percent of income is received by the lowest 20 percent of families. Point *B* shows that 13.9 percent of income is received by the lowest 40 percent of families. These two points and the others in the figure come from Table 14.4. In addition, the 45-degree line shows perfect equality, and the solid lines along the horizontal and right-hand vertical axes show perfect inequality. The shaded area between the 45-degree line and the Lorenz curve is a measure of inequality. The ratio of this area to the area of the triangle below the 45-degree line is the Gini coefficient. The Gini coefficient for 2000 is .462.

45-degree line that it would consist of a straight line on the horizontal axis up to 100 and then a vertical line. For example, 99.9 percent of the households would earn zero percent of the income. Only when the richest person is included do we get 100 percent of the income.

Gini coefficient: an index of income inequality ranging between 0 (for perfect equality) and 1 (for absolute inequality); it is defined as the ratio of the area between the Lorenz curve and the perfect equality line to the area between the lines of perfect equality and perfect inequality.

The **Gini coefficient** is a useful numerical measure of how bowed out the Lorenz curve is. It is defined as the ratio of the area of the gap between the 45-degree line and the Lorenz curve to the area between the lines of perfect equality and perfect inequality. The Gini coefficient can range between 0 and 1. It has a value of zero if the area between the diagonal line and the Lorenz curve is zero. Thus, when the Gini coefficient is zero, we have perfect equality. The Gini coefficient would be 1 if only one person earned all the income in the economy.

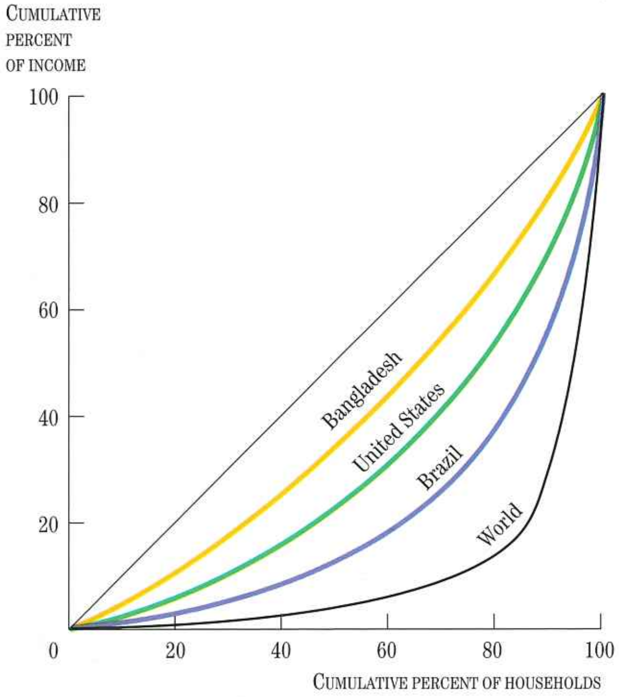
Economics in Action

Income Distribution Around the World

Lorenz curves can be calculated for different countries or groups of countries. For most European countries, the Lorenz curve is closer to equality than it is for the United States. Canada, Australia, and the United Kingdom have Lorenz curves very similar to that of the United States.

However, income distribution varies much more when we look beyond the developed countries. As the figure shows, Bangladesh, a very poor country, has a more equal income distribution than the United States. Brazil, a middle-income country that is also much poorer than the United States, has a much less equal income distribution. Among individual countries, Bangladesh and Brazil are close to the extremes: 60 percent of the population receives 40 percent of the income in Bangladesh and 19 percent of the income in Brazil, compared to 29.8 percent in the United States.

Income distribution for the world as a whole is far more unequal than that for any one country because the very poor in some countries are combined with the very rich in other countries. For example, when West Germany united with East Germany to form one country, the income distribution became more unequal for the unified country as a whole than it had been for either country before unification. The Lorenz curve for the world as a whole—as illustrated in the figure—shows far greater inequality than the curve for any one country: 60 percent of the world's population receives only 5 percent of the income.



Note: World curve computed from population data for low-, lower-middle-, upper-middle-, and high-income countries.

Figure 14.11 shows how the Gini coefficient has changed in the United States over the last 50 years. The Gini coefficient has varied within a narrow range, from .32 to .46. The most notable feature of the trend in the Gini coefficient in Figure 14.11 is the decline after World War II until around 1970 and the subsequent increase. It is clear that in recent years income inequality has increased. Higher earnings of skilled and educated workers relative to the less skilled and less educated may partly explain this change in income distribution. But the reason for these changes in income inequality is still a major unsettled question for economists.

It is important to note, however, that an increase in income inequality, as in Figure 14.11, does not necessarily mean that the rich got richer and the poor got poorer. For example, if one looks at average income in each quintile, one finds an increase for all groups from the 1970s to the 1990s, even after adjusting for inflation. However, average income in the top quintile increased by a larger percentage amount than average income in the bottom quintile.

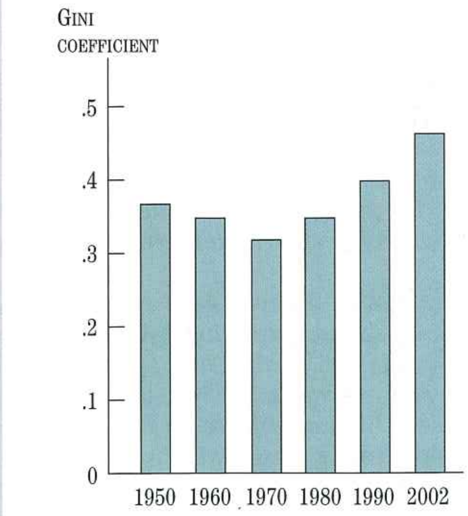


Figure 14.11
Changes in Income Inequality: The U.S. Gini Coefficient
The Gini coefficient is large when there is more inequality, as measured by the Lorenz curve in Figure 14.10. Thus, by this measure, inequality fell in the United States from 1950 to 1970 but increased from 1970 to 2002.

■ **Income Mobility and Longer-Term Income Inequality.** In interpreting income distribution statistics, it is important to recognize that the quintiles do not refer to the same people as the years go by. People move from quintile to quintile. People who are in the top quintile in one year may be in the bottom quintile the next year. And people who are in the bottom quintile one year may be in the top quintile the next year. In its 2003 Survey of Income and Program Participation, the U.S. Census Bureau reported that 38 percent of those households in the lowest income quintile in 1996 were in a higher income quintile in 1999.

Distinguishing between income in any one year and income over several years is important. In a typical life span, people usually earn less when they are young than when they are middle-aged. As they grow older and become more experienced, their wages and salaries increase. When people retire, their income usually declines again. Thus, even if everyone had the exact same lifetime income, one would see inequality in the income distribution every year. Middle-aged people would be relatively rich, while young and old people would be relatively poor.

More generally, many people seem to move around the income distribution from year to year. Some undergo hardship, such as a layoff or a permanent loss of job because of a change in the economy. Even if they are eventually rehired, in the short run they fall to the lower end of the income distribution when they are unemployed. On the other hand, some people do well and move quickly to the top of the income distribution.

How significant is income mobility? Economic research shows that about two-thirds of the people in any one quintile move to another over a 10-year period. About half of those in the top quintile move to a lower quintile, and about half of those in the bottom quintile move to a higher quintile. This degree of mobility has not changed in the last 20 years.

■ **Changing Composition of Households.** The formation or splitting up of households can also affect the distribution of income. For example, if two individuals who were living separately form a household, the household income doubles. Households splitting apart can also alter the income distribution drastically. If one adult leaves the family, perhaps because of divorce or desertion, and the other one stays home with the children, the income of the household declines substantially. In 2003, 28 percent of families with single mothers were found to be living below the poverty line, as compared to a rate of 10 percent for all families.

It appears that the splitting apart of households has had an impact on income distribution in the United States. According to some estimates, if household composition had not changed in the United States in the last 20 years, there would have been only half as great an increase in inequality as measured by the Gini coefficient.

Distribution of Income versus Distribution of Wealth. Another factor to keep in mind when interpreting data on the income distribution is the distinction between *income* and *wealth*. Your annual income is what you earn each year. Your wealth, or your net worth, is all you own minus what you owe others. Wealth changes over a person's lifetime even more than income does as people save for retirement. For example, a young person who has just graduated may have little wealth or, with a college loan still to pay off, may have negative net worth. However, by saving a bit each year—perhaps through a retirement plan at work—the person gradually accumulates wealth. By the time retirement age is reached, the person may have a sizable retirement fund, and thus be relatively wealthy.

A survey of about 3,000 families in the United States in 2001 found that the top 10 percent of households held about 72 percent of the net worth in the United States. Although such surveys are not as accurate as the regular monthly surveys of 60,000 households on which our information about income distribution is based, it is clear that the distribution of wealth is less equal than the distribution of income. About one-third of the net worth in this survey was in the form of net worth held in small businesses.

The Poor and the Poverty Rate

Many believe the main purpose of government redistribution of income is to help the poor. For this reason, the term *social safety net* is sometimes used for an income redistribution system; the idea is that these programs try to prevent those who were born poor or who have become poor from falling too far down in income and therefore in nutrition, health, and general well-being.

Poverty can be observed virtually everywhere. The poor are visible in the blighted sections of cities and in remote rural areas. Almost everyone has seen the serious problems of the homeless in cities of the United States. CNN has brought the agony of poverty around the world to our TV screens.

poverty rate: the percentage of people living below the poverty line.

poverty line: an estimate of the minimum amount of annual income required for a family to avoid severe economic hardship.

To gauge the success or failure of government policies to alleviate poverty, economists have developed quantitative measures of poverty. The **poverty rate** is the percentage of people who live in poverty. To calculate the poverty rate, one needs to define what it means to live in poverty. In the United States, poverty is usually quantitatively defined by a **poverty line,** an estimate of the minimal amount of annual income a family needs in order to avoid severe economic hardship. The poverty line in the United States is based on a survey showing that families spend, on average, one-third of their income on food. The poverty line is thus obtained by multiplying by 3 the Department of Agriculture's estimate of the amount of money needed to purchase a low-cost nutritionally adequate amount of food. In addition, adjustments are made for the size of the family. Table 14.5 shows the poverty line for several different family sizes. Since the 1960s, when it was first developed, the poverty line has been increased to adjust for inflation.

Table 14.5
The Poverty Line in the United States, 2004

Family Size	Poverty Line
Unrelated individuals	$ 9,827
Two persons	$12,649
Three persons	$14,776
Four persons	$19,484
Seven persons	$31,096

Source: U.S. Census Bureau.

Using the poverty line and data on the income distribution, one can determine the number of people who live in poverty and the poverty rate. The overall poverty rates in the United States have varied over the last 50 years. The overall poverty rate for families declined sharply from 18 percent in 1960 to 10 percent in 1970, but has remained relatively constant since 1970. In 2004 the poverty rate for families was 10 percent, while the poverty rate for all individuals was 12.7 percent, or 37 million people.

There are important trends in poverty for different groups in the population. For example, the percentage of children who live in poverty rose in the 1970s, 1980s, and early 1990s. During the same time period, the poverty rate for the elderly has declined, and this has held down the overall poverty rate. In 1993, when child poverty was at a peak, 22.7 percent of children were living in poverty; at the same time, the poverty rate for the elderly was 12.2 percent. The dramatic decline in the poverty rate for the elderly (from 24.6 percent in 1970), is largely attributed to the change in Social Security benefits during that time period—along with better retirement benefits.

The increase in poverty for children is troublesome, and it is difficult to explain. Some of it may have to do with the increase in single-headed households with children, which are usually poorer than two-adult households. Poverty rates in households with a single head and at least one child have ranged between 35 and 40 percent in the last 20 years—three times the overall poverty rate. However, since 1993, the poverty rate for children has made some progress, declining from 22.7 percent to 16.2 percent in 2000. Since then, it has risen slightly to 17.8 percent in 2004.

Effects of Taxes and Transfers on Income Distribution and Poverty

We have seen that two goals of the tax and transfer system are to redistribute income and reduce poverty. How successful is this redistribution effort? Estimates by the U.S. Census Bureau indicate that the tax and transfer system reduces the poverty rate by about 10 percentage points: Without the tax and transfer system, the Census Bureau estimates that the poverty rate would be 20 percent rather than 10 percent. Those in the second quintile have their average income increased by an average of about $4,000 as a result of the tax and transfer system. Those in the uppermost part of the income distribution have their average income reduced by about $22,000 as a result of the tax and transfer system.

To be sure, these estimates ignore any of the incentive effects mentioned earlier, such as the reduced work incentives that might result from the tax and transfer system. And they ignore any possible response of private efforts to redistribute income—such as charities—that might occur as a result of changes in the government's role.

REVIEW

- The distribution of income can be measured by the percentage of income earned by quintiles of households or families. The Lorenz curve and the Gini coefficient are computed from this distribution.
- The poverty rate is a quantitative measure of the amount of poverty in the United States.
- The distribution of income has become more unequal in the United States since the early 1970s. The change is partly due to a growing dispersion of wages between unskilled workers with little education and highly skilled workers with more education.

- Over the last three decades, the poverty rate among children has increased while the poverty rate for the elderly has declined.
- Estimates indicate that the tax and transfer system currently makes the poverty rate lower than it would otherwise be.
- Nevertheless, the increase in poverty among children has raised serious concerns about the tax and transfer system in the United States.

Conclusion

In a democracy, the amount of government redistribution of income is decided by the people and their representatives. A majority seem to want some redistribution of income, but there is debate about how much the government should do.

Why doesn't a democracy lead to much more redistribution? After all, 60 percent of the people, according to Table 14.4, receive only 30 percent of the income. Since 60 percent of the voting population is enough to win an election, this 60 percent could vote to redistribute income much further. Why hasn't it?

There are probably a number of reasons. First, there is the tradeoff between equality and efficiency stressed in this chapter. People realize that taking away incentives to work will reduce the size of the pie for everyone.

Another reason is that most of us believe that people should be rewarded for their work. Just as we can think of a fair income distribution, we can think about a fair reward system. If some students want to work hard in high school so that they can attend college, why shouldn't they get the additional income that comes from that?

There is also the connection between personal freedom and economic freedom. Government involvement in income distribution means government involvement in people's lives. Those who cherish the idea of personal freedom worry about a system that takes a large amount of income from people who work.

Finally, much income redistribution occurs through the private sector—private charities and churches. The distribution of food and the provision of health care have long been supported by nongovernment organizations. In times of floods or earthquakes, it is common for people to volunteer to help those in distress. Private charity has certain advantages over government. Individuals become more personally involved if they perform a public service, whether volunteering at a soup kitchen or tutoring at an elementary school. But incentives for redistribution through private charity may be too small. People may give less to a charitable organization if they believe others are not giving. Thus, the private sector may not be sufficient.

KEY POINTS

1. The government in modern democracies plays a major role in trying to help the poor and provide a more equal income distribution.
2. Taxes are needed to pay for transfers and other government spending. In the United States, the personal income tax and the payroll tax are by far the most significant sources of tax revenue at the federal level. Sales taxes and property taxes play a significant role at the local level.
3. Taxes cause inefficiencies, as measured by deadweight loss, because taxes reduce the amount of the economic activity being taxed—whether it is the production of a good or the labor of workers.
4. The incidence of a tax depends on the price elasticity of supply and demand. The deadweight loss from taxes on goods with low price elasticities is relatively small.

5. Transfer payments are classified into means-tested programs—such as welfare and food stamps—and social insurance programs—such as social security and unemployment insurance.
6. Transfer payments can cause inefficiency as a result of disincentives to work or the incentive for families to split up.
7. There is a tradeoff between equality and efficiency. Tax reform and welfare reform try to improve incentives and reduce inefficiency.
8. The distribution of income has grown more unequal in recent years.
9. Poverty among children has increased, while poverty among the elderly has declined in recent years.
10. The tax and transfer system has reduced income inequality and lowered poverty rates, but there is much room for improvement and reform.

KEY TERMS

personal income tax
taxable income
tax bracket
marginal tax rate
average tax rate
progressive tax
regressive tax
proportional tax
flat tax
payroll tax
corporate income tax
excise tax
sales tax
property tax
tax incidence
tax revenue
ability-to-pay principle
transfer payment
means-tested transfer
social insurance transfer
family support programs
Medicaid
supplemental security income (SSI)
food stamp program
Head Start
housing assistance programs
earned income tax credit (EITC)
social security
Medicare
unemployment insurance
mandated benefits
Current Population Survey
quintiles
Lorenz curve
Gini coefficient
poverty rate
poverty line

QUESTIONS FOR REVIEW

1. What are the two largest sources of tax revenue for the federal government?
2. What is the difference between income and taxable income?
3. Why is there a deadweight loss from the personal income tax?
4. Why are the effects of a payroll tax the same whether the employer or the worker pays it?
5. Why is deadweight loss from a payroll tax small when the elasticity of labor supply is small?
6. What is the difference between a marginal tax rate and an average tax rate? Why is the former more important for incentives?
7. What causes the tradeoff between equality and efficiency?
8. In what way do both welfare reform and tax reform focus on incentives?
9. What was the primary goal of the welfare reform law passed in 1996?
10. How is the distribution of income measured by the Lorenz curve and the Gini coefficient?
11. Why are income mobility and lifetime income important for interpreting the income distribution statistics?

PROBLEMS

1. The following table gives the income distribution in Brazil and in Australia. Draw the Lorenz curve for each. Which country has the larger Gini coefficient?

Quintile	Percent of Income in Brazil	Percent of Income in Australia
Bottom 20 percent	2.4	4.4
Second 20 percent	5.7	11.1
Third 20 percent	10.7	17.5
Fourth 20 percent	18.6	24.8
Top 20 percent	62.6	42.2

2. Suppose the government decides to increase the payroll tax paid by employers. If the labor supply curve has a low elasticity, what will happen to the workers' wages? Who actually bears the burden of the tax, the workers or the firms? Would it be different if the labor supply had a high elasticity?
3. The table on the next page gives hours worked and the welfare payment received.
   a. Calculate the missing data in the table, given that the hourly wage is $5 and total income is the sum of the wage payment and the welfare payment.

Hours Worked	Wage Payment	Welfare Payment	Total Income
0		10,000	
500		8,000	
1,000		6,000	
1,500		4,000	
2,000		2,000	

b. Draw a graph that shows how much total income a worker earns with and without this welfare program. Put the number of hours worked on the horizontal axis and total income on the vertical axis.

c. What is the increase in total income for each additional hour worked without any welfare program? Compare it with the increase in total income for each additional hour worked under the welfare program.

d. How could the welfare program be changed to increase the incentive to work without reducing total income for a full-time worker (40 hours per week, 50 weeks per year, $5 per hour) below $12,000, which is already below the poverty line for a family of four?

4. Suppose that the labor demand curve is perfectly flat. What is the impact on a typical worker's hourly wage if the government increases the payroll tax paid by employers by 10 percent of the wage? Show what happens in a labor supply and labor demand graph like Figure 14.6. Why does the slope of the labor supply curve not affect your answer?

5. Given the data in the table below, draw the Lorenz curve before and after the proposed tax. Is this tax progressive or regressive?

Quintile	Percent of Income Without Tax	Percent of Income with Tax
Bottom 20 percent	6	7
Second 20 percent	9	12
Third 20 percent	18	20
Fourth 20 percent	25	25
Top 20 percent	42	36

6. The Family Leave Act is a federal law that requires employers to give unpaid leave to employees to care for a newborn or a sick relative. Show how the Family Leave Act affects the supply and demand for labor. According to this model, what will happen to wages and employment compared to the prelaw situation?

7. Many states do not tax food items because that kind of tax is considered regressive. Explain. California tried to impose a "snack" tax—one applying only to what the legislators thought was junk food. Suppose snack food has a higher elasticity of demand than nonsnack food. Draw a supply and demand diagram to explain which tax—on snack food or nonsnack food—will cause the price to rise more. Which will have a greater deadweight loss?

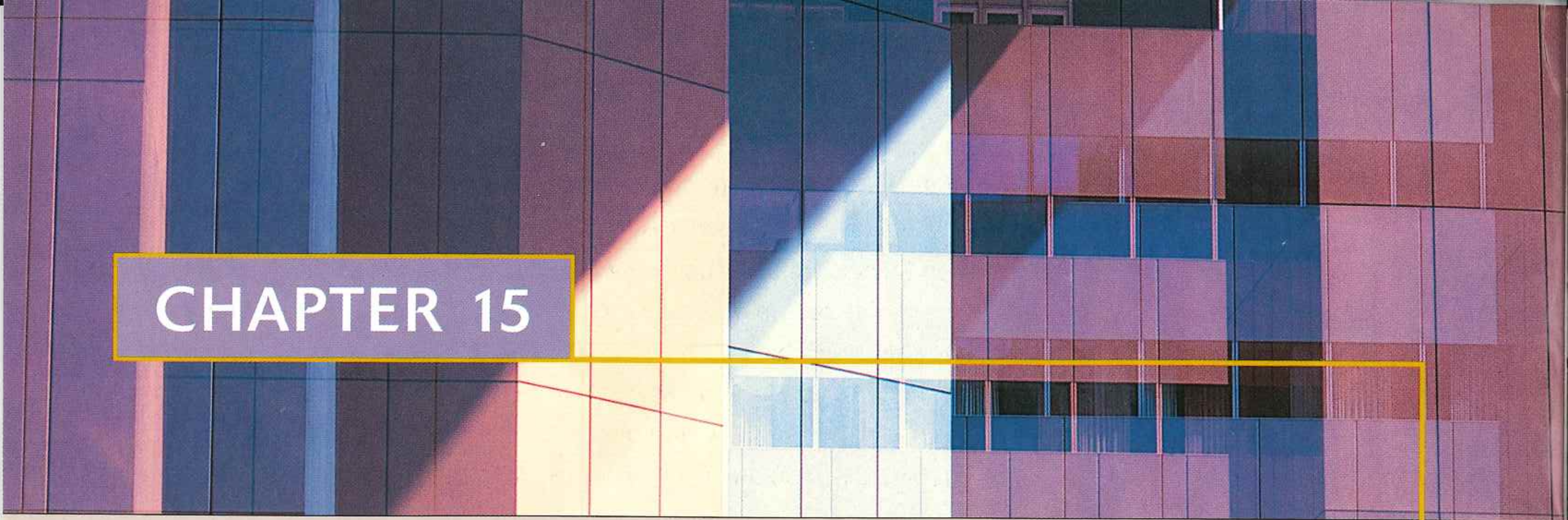
8. Some economists argue that we should use more progressive taxes, while others claim that we should adopt a flat tax. List some reasons for and against using progressive taxes.

9. Suppose the government is trying to decide between putting a sales tax on luxuries, which usually have very high demand elasticities, or on gasoline. Which tax will have a bigger effect on the market price? Which tax will cause the quantity traded in the market to decline more? Draw a diagram to explain.

10. Suppose Fred, a bookkeeper, had taxable income of $21,000 last year. His doctor, Celia, had taxable income of $140,000 last year. Use the tax rate schedule in Figure 14.2 to figure out how much each owes in income taxes. What are their marginal tax rates? What are their average tax rates? (Assume that both are single.)

11. Analyze the distribution of income, using the household incomes in the following table. Rank the families by income. Compute the percentage of total income going to the poorest 20 percent of the families, the second 20 percent, and the richest 20 percent. Draw a Lorenz curve for the income distribution of these 10 families. Is their distribution more equal or less equal than that of the population of the United States as a whole?

Family	Income
Jones	$ 12,000
Pavlov	$100,000
Cohen	$ 24,000
Baker	$ 87,000
Dixon	$ 66,000
Sun	$ 72,000
Tanaka	$ 18,000
Bernardo	$ 45,000
Smith	$ 28,000
Lopez	$ 33,000



CHAPTER 15

Public Goods, Externalities, and Government Behavior

Economists who have worked at the President's Council of Economic Advisers in Washington are frequently asked what they did there. One answer—given recently by an economist who worked on President Kennedy's Council way back in the early 1960s—is timeless; it still applies today, and it will undoubtedly apply in the future. The economist was Robert Solow,[1] who later won a Nobel Prize in economics.

Solow put it this way: "On any given day in the executive branch, there are more meetings than Heinz has varieties. At a very large proportion of these meetings, the representative of some agency or some interest will be trying to sell a harebrained economic proposal. I am exaggerating a little. Not every one of these ideas is crazy. Most of them are just bad: either impractical, inefficient, excessively costly, likely to be accompanied by undesirable side effects, or just misguided—unlikely to accomplish their stated purpose. Someone has to knock those proposals down. . . . That is where the Council's comparative advantage lies." Solow emphasized that he had good people and good arguments to work with, but that "does not mean that we won all the battles; we lost at least as many

[1] Robert M. Solow, "It Ain't the Things You Don't Know That Hurt You, It's the Things You Know That Ain't So," *American Economic Review*, May 1997, pp. 107–108.

as we won. The race is not always won by the best arguments, not in political life anyway. But we always felt we had a chance and we kept trying."

The purpose of this chapter is to examine two important concepts—public goods and externalities—that economists on the President's Council and elsewhere use to determine whether a proposal for government action is bad or good. We also show how cost-benefit analysis can be used to help determine the correct course of action. Finally, we examine different models of government behavior to understand why "in political life" the best economic arguments do not always win. We will see that politicians are influenced by incentives as much as anyone else.

Public Goods

Table 15.1 shows the range of goods and services produced by all governments in the United States: the federal government, 50 state governments, and 87,453 local governments (counties, cities, towns, and school districts). Education is by far the largest in terms of employment, followed by health and hospital services, national defense, police, the postal service, and highways. (The figures for national defense include only civilian workers; if Table 15.1 included those serving in the armed forces, national defense would be second on the list.) The other categories, from the judicial and legal system (federal, state, and county courts) to parks and recreation, are each significant but small relative to the total.

Observe also the types of goods and services that are not on the list because they are produced by the private sector. Manufacturing, mining, retail trade, wholesale trade, hotel services, and motion picture production are some of the items largely left to the private sector. Note also that for all the goods and services on the list in Table 15.1, the private sector provides at least some of the production. There are 6 million workers in the private health-care sector, for example, compared to 2 million in government health care. The private sector is also involved in mail delivery, education, garbage collection, and even fire protection (volunteer fire departments).

Table 15.1
Types of Goods and Services Produced by Government in the United States

Type of Good or Service	Employment as a Percent of Total Government
Education	45
Health and hospitals	10
National defense (civilian)	5
Police protection	5
Postal service	4
Highways	3
Judicial and legal	2
Parks and recreation	2
Fire protection	2
Sanitation, sewage	1
All other	21

Nonrivalry and Nonexcludability

Why is it necessary for governments to produce *any* goods and services? The concept of a public good helps us answer the question. A **public good** is a good or service that has two characteristics: *nonrivalry in consumption* and *nonexcludability.*

Nonrivalry in consumption means that more consumption of a good by one person does not imply less consumption of it by another person. For example, if you breathe more clean air by jogging rather than watching television, there is no less clean air for others to breathe. Or when a new baby is born, the baby immediately benefits from national defense without anyone else having to give up the benefits of national defense. Once a country's national defense—the military personnel, the strategic alliances, the missile defense system—is in place, the whole nation enjoys the security simultaneously; the total benefit is the sum of the benefits of every person. Clean air and national defense are examples of goods with nonrivalry. In contrast, for most goods, there is rivalry in consumption. For example, if you consume

public good: a good or service that has two characteristics: nonrivalry in consumption and nonexcludability.

nonrivalry: the situation in which increased consumption of a good by one person does not decrease the amount available for consumption by others.

In the wake of Hurricane Katrina A Chinook helicopter drops sand bags to plug a canal levee break in the Gentilly neighborhood of New Orleans, Louisiana on September 11, 2005. Levees are a public good, having characteristics of both non-rivalry and non-excludability.

more french fries, then either someone else must consume fewer french fries or more french fries must be produced. But for a good with nonrivalry in consumption, everybody can consume more if they want to. There is a collective aspect to the good.

nonexcludability: the situation in which no one can be excluded from consuming a good.

Nonexcludability means that one cannot exclude people from consuming the good. For example, people cannot be excluded from the benefits of national defense. In contrast, most goods have the characteristic of excludability. If you do not pay for the french fries that you ordered at the drive-through window, the server will not give them to you.

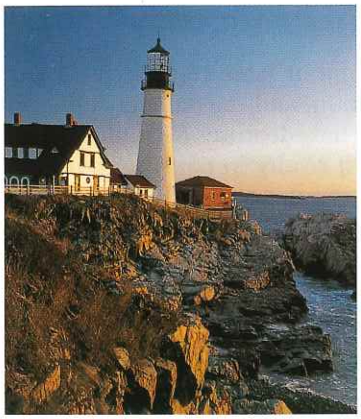
A public good is a good or service that has nonrivalry in consumption and nonexcludability. In contrast, a *private* good has excludability and rivalry.

Free Riders: A Difficulty for the Private Sector

free-rider problem: a problem arising in the case of public goods because those who do not contribute to the costs of providing the public good cannot be excluded from the benefits of the good.

Goods that have nonrivalry in consumption and nonexcludability create a **free-rider problem:** People can enjoy the good or service without reducing others' enjoyment even if they do not pay. To understand this concept, imagine that you bought a huge bus for the purpose of transporting students around town and collecting a little money for your service. But suppose the bus had a broken rear door that allowed people to get on and off without paying and without interfering with other people's travel. In that situation, you would have free riders. If you could not fix the door or do something else to exclude the free riders, you would not be in the transportation business long, because without fares, you would have losses.

National defense is like the huge bus with the broken rear door. You cannot exclude people from enjoying it, even if they do not pay, and one person's security does not reduce the security of others. It is clear that a private firm will have difficulty producing and selling national defense to the people of a country. For this reason, a collective action of government to provide this defense, requiring that people pay for it with taxes, is necessary. Similar actions are taken with other public goods such as police protection, fire protection, and the judicial system. That clean air has the property of a public good explains why government is involved in its "production." In this case, the government might help produce clean air by prohibiting the burning of



The Classic Lighthouse Example
How can the free-rider problem be avoided without government providing the service?

leaves or the using of backyard barbecues. We will return to the government's role in air quality when we discuss the concept of externality.

Information also has the features of a public good. Everyone can benefit from information that a hurricane is on the way; there is no rivalry in consuming this information. Information about the state of the economy can also benefit everyone. For this reason, such information has been largely supplied by governments. In the United States, the Department of Commerce collects and distributes information about the economy, and the U.S. Weather Service collects information about the weather.

Avoiding Free-Rider Problems

Not all public goods are provided by the government. When they are not, some other means of dealing with the free-rider problem is needed. A classic example used by economists to explore the nature of public goods is the lighthouse that warns ships of nearby rocks and prevents them from running aground. A lighthouse has the feature of nonrivalry. If one ship enjoys the benefit of the light and goes safely by, this does not mean that another ship cannot go by. There is no rivalry in the consumption of the light provided by the lighthouse. Similarly, it is impossible to exclude ships from using the lighthouse because any ship can benefit from the light it projects.

However, lighthouse services are not always provided by the government. Early lighthouses were built by associations of shippers, who charged fees to the ships in nearby ports. This system worked well because the fees could be collected from most shippers as they entered nearby ports. The free-rider problem was avoided, so general tax revenue to pay for the lighthouse was not needed.

When the users of a government-provided service are charged for its use (some excludability is needed), the charge is called a **user fee.** In recent years, user fees have become more common in many government-provided services, including the national parks. The aim is to target the payments more closely to the users of the goods and services.

user fee: a fee charged for the use of a good normally provided by the government.

Although police services are almost always provided by government, there are many examples of security services provided by private firms. For example, a business firm may hire a guard to watch its premises. In these cases, the service is focused at a particular group, and excludability is possible, and so the free-rider problem can be avoided.

New Technology

Modern technology is constantly changing the degree to which there is nonrivalry and nonexcludability for particular goods. When radio and television were invented, it became clear that once a radio or television program was broadcast, it was possible for anyone to tune in to the broadcast. A radio or television broadcast has both characteristics of a public good. But private firms have provided the vast majority of radio and television broadcasting services in the United States. The free-rider problem was partially avoided by using advertising to pay for the service. Paying directly would be impossible because of the inability to exclude individuals who do not pay.

More recently, technology is changing the public good features of television. Cable TV and the ability to scramble signals for those who use satellites to obtain their television signals have reduced the problem of nonexcludability. If one does not pay a cable television bill, the service can be turned off. If one does not pay the satellite fee, the signals can be scrambled so that reception is impossible. Thus, it is now common to see cable television stations delivering specialized programming to small audiences that pay extra for the special service.

Public Goods and Actual Government Production

If we look at the types of goods produced by government in Table 15.1, we see many public goods, such as national defense, police protection, and the judicial and legal system. However, many of the goods in the list do not have features of public goods. Postal delivery, for example, is a service that has both rivalry in consumption and excludability. If you do not put a stamp on your letter, it is not delivered, and there is certainly rivalry in the consumption of a postal delivery worker's time. In principle, education also is characterized by rivalry in consumption and excludability. For a given-sized school, additional students reduce the education of other students, and it is technologically possible to exclude people. Although there are other reasons why the government might be involved in the production of these goods, it is important to note that the production of a good by the government does not make that good a public good. In centrally planned economies, the government produces virtually everything, private goods as well as public goods. The economist's definition of a public good is specific and is useful for determining when the government should produce something and when it is better left to the market.

Cost-Benefit Analysis

Suppose it is decided that a good or service is a public good and that if it is to be produced at all, government should provide it. Should the good be produced? How much of the good should be produced? Such decisions are ultimately made by voters and elected officials after much political debate. Some economic analysis of the costs and benefits of the goods and services should inform the participants in this debate. Balancing the costs and benefits of a good or service is called **cost-benefit analysis.**

cost-benefit analysis: an appraisal of a project based on the costs and benefits derived from it.

Marginal Cost and Marginal Benefit. To determine the quantity of a government-provided service that should be produced, the marginal cost and marginal benefit of the service should be considered. In the case of police services, for example, a decision about whether to increase the size of the police force should consider both the marginal benefit to the people in the city—the reduction in the loss of life and property caused by crime, the increased enjoyment from a secure environment, and safer schools—and the marginal cost—the increased payroll for the police. If the marginal benefit of more police is greater than the marginal cost of more police, then the police force should be increased. The optimal size of the police force should be such that the marginal cost of more police is equal to the marginal benefit of more police.

Measuring the costs of producing government-provided services is not difficult because government workers' wages and materials used in production have explicit dollar values.

But measuring the benefits of government-provided services is much more difficult. How do we measure how much people value greater security in their community? How do we value a reduction in violence at schools or a reduced murder rate? Public opinion polls in which people are asked how much they would be willing to pay are a possibility. For example, people can be asked in surveys how much they would be willing to pay for more police in an area. Such estimates of willingness to pay are called **contingent valuations** because they give the value contingent on the public good's existing and the person's having to pay for it. Some economists think that contingent valuation is not reliable if people do not actually have to pay for the good or service. People may not give a good estimate of their true willingness to pay.

contingent valuation: an estimation of the willingness to pay for a project on the part of consumers who may benefit from the project.

REVIEW

- Public goods have two characteristics: nonrivalry, which means that greater consumption for one person does not mean less consumption for someone else, and nonexcludability, which means that it is not possible to exclude those who do not pay for the good.
- Public goods have a free-rider problem, which means that government production is frequently necessary.
- The private sector must deal with the free-rider problem if such goods are to be produced in the market.
- In deciding how much of a public good should be provided, cost-benefit analysis can be used. In deciding how large a police force should be, for example, the quantity of police services produced should be such that the marginal benefit of additional police equals marginal cost.

Externalities: From the Environment to Education

We have seen that the existence of public goods provides an economic rationale for government involvement in the production of certain goods and services. Another rationale for government involvement in production is a market failure known as an externality. An **externality** occurs when the costs of producing a good or the benefits from consuming a good spill over to individuals who are not producing or consuming the good. The production of goods that cause pollution is the classic example of an externality. For example, when a coal-fired electric utility plant produces energy, it emits smoke that contains carbon dioxide, sulfur dioxide, and other pollutants into the air. These pollutants can make life miserable for people breathing the air and cause serious health concerns. Similarly, automobiles emit pollutants and reduce the quality of life for people in areas where cars are driven. Those who drive cars add a cost to others. These are examples of **negative externalities** because they have a negative effect—a cost—on the well-being of others. A **positive externality** occurs when a positive effect—a benefit—from producing or consuming a good spills over to others. For example, you might benefit if your neighbor plants a beautiful garden that is visible from your house or apartment. Let us first look at the effects of negative externalities and then consider positive externalities.

externality: the situation in which the costs of producing or the benefits of consuming a good spill over onto those who are not producing or consuming the good.

negative externality: the situation in which costs spill over onto someone not involved in producing or consuming the good.

positive externality: the situation in which benefits spill over onto someone not involved in producing or consuming the good.

Negative Externalities

In the case of negative externalities, a competitive market does not generate the efficient amount of production. The quantity produced is greater than the efficient quantity. For example, too much air-polluting electrical energy may be produced. The reason is that producers do not take into account the external costs when they calculate their costs of production. If they did take these costs into account, they would produce less.

The reason why competitive markets are not efficient in the case of negative externalities can be illustrated using the supply and demand curves. For example, consider an example of a negative externality due to pollution caused by the

Oil Spill: A Negative Externality
The oil spilled into the ocean by this sinking oil tanker is an example of a negative externality: The production of goods or services (transportation of oil) by a firm raises costs or reduces benefits to people (the oil spill).

production of electricity. A negative externality occurs because the production of electricity raises pollution costs to other firms or individuals. The electrical utility plant pollutes the air and adds costs greater than the costs perceived by the electrical utility. The externality makes the marginal cost as perceived by the private firm, which we now call the **marginal private cost,** less than the true marginal cost that is incurred by society, which we call the **marginal social cost.** Marginal social cost is the sum of the firm's marginal private cost and the increase in external costs to society as more is produced. The marginal external cost is the change in external costs as more is produced. That is,

marginal private cost: the marginal cost of production as viewed by the private firm or individual.

marginal social cost: the marginal cost of production as viewed by society as a whole.

Marginal social cost = marginal private cost + marginal external cost

We illustrate this in Figure 15.1 by drawing a marginal private cost curve below the marginal social cost curve. We use the term *marginal private cost* to refer to what we have thus far called marginal cost in order to distinguish it from marginal social cost. Recall that adding up all the marginal (private) cost curves for the firms in a market gives the market supply curve, as labeled in the diagram.

Figure 15.1 also shows the marginal benefit to consumers from using the product, in this case electrical energy. This is the market demand curve for electricity. According to the supply and demand model, the interaction of firms and consumers in the market will result in a situation in which the marginal cost of production—the marginal private cost—equals marginal benefit. This situation occurs at the market equilibrium, where the quantity supplied equals the quantity demanded. The resulting quantity produced is indicated by point *B* in Figure 15.1.

However, at this amount of production, the marginal benefit of production is less than the marginal *social* cost of production. Marginal benefit equals marginal private cost but is less than marginal social cost. Only at point *A* in the figure is marginal benefit equal to marginal social cost. Thus, point *A* represents the efficient level of production. Because of the externality, too much is produced. Firms produce too much because they do not incur the external costs. There is a deadweight loss, as shown in Figure 15.1. Consumer surplus plus producer surplus is not maximized.

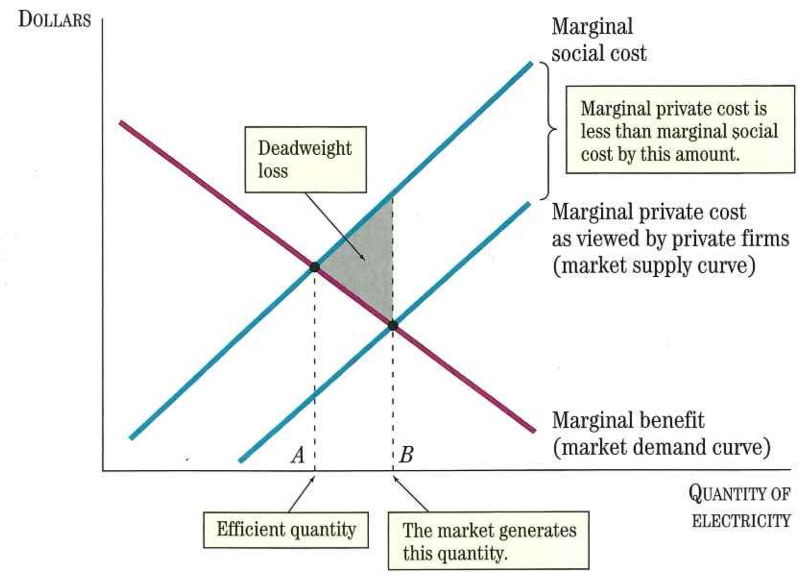


Figure 15.1
Illustration of a Typical Negative Externality
Because production of the good creates costs external to the firm (for example, pollution), the marginal social cost is greater than the marginal private cost to the firm. Thus, the equilibrium quantity that emerges from a competitive market is too large: Marginal benefit is less than marginal social cost.

Positive Externalities

A positive externality occurs when the activity of one person makes another person better off, either reducing costs or increasing benefits. Let us examine what happens when a positive externality raises social benefits above private benefits. For example, increased earnings are a benefit from attending high school, college, graduate school, or continuing education. But the education also benefits society. The greater education that these individuals receive is spread to others. Going to school and learning to read and write makes people better citizens. Learning about hygiene and becoming health conscious puts less of a burden on the public health system.

Another example of a good with a positive externality is research. Firms that engage in research get some of the benefits of that research through the products that they can sell—maybe novel products. But in many cases the research spreads, and other people can take advantage of it as well. Research that spills over to other industries or other individuals is an externality. The benefit from the research expenditures goes beyond the individual; it creates inefficiencies just as negative externalities do.

To show how positive externalities affect the quantity produced in a competitive market, we need to look at the supply and demand curves. The externality makes the marginal benefit as perceived by the consumer, which we now call the **marginal private benefit,** less than the true benefit to society, which we call the **marginal social benefit.** With a positive externality, the marginal social benefit is greater than the

marginal private benefit: the marginal benefit from consumption of a good as viewed by a private individual.

marginal social benefit: the marginal benefit from consumption of a good from the viewpoint of society as a whole.

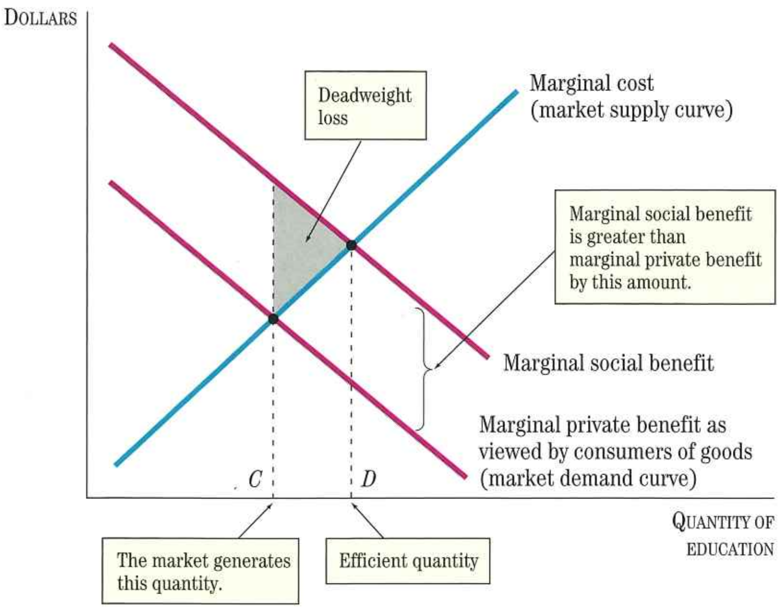


Figure 15.2
Illustration of a Typical Positive Externality
Because consumption of the good (for example, education) gives benefits to others, the marginal social benefit is greater than the marginal private benefit. Hence, the equilibrium quantity that emerges from a competitive market is too low.

marginal private benefit because there is a marginal external benefit from more consumption. That is,

Marginal social benefit = marginal private benefit + marginal external benefit

Figure 15.2 shows the impact of this difference between marginal social benefit and marginal private benefit. Suppose that Figure 15.2 refers to the market for education. Then the quantity of education is on the horizontal axis. The marginal social benefit curve is above the marginal private benefit curve in the figure. The marginal private benefit curve for consumers is the market demand curve. Consider the equilibrium quantity at point *C*, where the quantity supplied equals the quantity demanded in Figure 15.2. The market results in a quantity produced that is less than the efficient quantity, which occurs when the marginal *social* benefit equals the marginal cost, as shown at point *D*. The quantity generated by the market is at a point where the marginal social benefit is greater than the marginal cost. Production and consumption of education are inefficient; the quantity of education is too low. Again, there is a deadweight loss due to externality, as shown in Figure 15.2.

Externalities Spread Across Borders

Externalities are sometimes international problems. Sulfur dioxide emissions from electrical utility plants are an externality whose international effects have received much attention. The sulfur dioxide travels high into the air and is then dispersed by winds across long distances. Rainfall then brings the sulfur dioxide back to earth in the form of acid rain, which lands on forests and lakes hundreds of miles away. In some cases, the acid rain occurs in countries different from the country in which the sulfur dioxide was first emitted. In North America, acid rain that results from burning fuel in the Midwest industrial centers may fall in Canada or upstate New York.

Global warming is another example of an externality with international dimension. When too much carbon dioxide accumulates in the earth's atmosphere, it prevents the sun's warmth from escaping out of the atmosphere, causing a greenhouse effect. Global warming is caused by the emission of carbon dioxide by firms and individuals but has effects all over the world.

REVIEW

- Externalities occur when the benefits or costs of producing and consuming spill over to others. Externalities cause the marginal private cost to be different from the marginal social cost, or the marginal private benefit to be different from the marginal social benefit.
- Externalities are a cause of market failure. Production of goods with negative externalities is more than the efficient amount. Production of goods with positive externalities is less than the efficient amount.
- Many externalities are global, occurring across borders, as when pollution emitted in one country has negative effects in other countries.

Remedies for Externalities

As the previous section shows, competitive markets do not generate an efficient level of production when externalities exist. What are some of the ways in which a society can alleviate problems caused by these externalities? In some cases, the solution has been for government to produce the good or service. In practice, elementary education is provided by governments all over the world with requirements that children attend school through a certain age. Education is by far the government-produced good or service with the most employment in the United States. But in most cases where externalities are present, production is left to the private sector, and government endeavors to influence the quantity produced. In fact, much of college education and some K–12 education is provided by the private sector in the United States.

How can production in the private sector be influenced by government so as to lead to a more efficient level of production of goods and services in the economy? We will see that the answer involves changing behavior so that the externalities are taken into account internally by firms and consumers. In other words, the challenge is to **internalize** the externalities.

internalize: the process of providing incentives so that externalities are taken into account internally by firms or consumers.

There are four alternative ways to bring about a more efficient level of production in the case of externalities. The first one discussed here, private remedies, does not require direct government intervention. The other three—command and control, taxes or subsidies, and tradable permits—do.

Private Remedies: Agreements Between the Affected Parties

private remedy: a procedure that eliminates or internalizes externalities without government action other than defining property rights.

In some cases, people, through **private remedies,** can eliminate externalities themselves without government assistance. A Nobel Prize winner in economics, Ronald Coase of the University of Chicago, pointed out this possibility in a paper published in 1960.

Consider the following simple example. Suppose that the externality relates to the production of two products: health care and candy. Suppose that a hospital is built next door to a large candy factory. Making candy requires noisy pounding and vibrating machinery. Unfortunately, the walls of the new hospital are thin. The loud candy machinery can be heard in the hospital. Thus, there is an externality that we might call noise pollution. It has a cost. It makes the hospital less effective; for example, it is difficult for the doctors to hear their patients' hearts through the stethoscopes.

What can be done? The city mayor could adopt a rule prohibiting loud noise near the hospital, but that would severely impinge on the candy making in the city. Or, because the hospital was built after the candy factory, the mayor could say, "Too bad, doctors; candy is important too." Alternatively, it might be better for the candy workers and doctors to work this externality out themselves. The supervisor of the candy workers could negotiate with the doctors. Perhaps the candy workers could agree to use the loud machines only during the afternoon, during which the doctors would take an extended break. Or perhaps a thick wall could be built between the buildings.

Thus, it is possible to resolve the externality by negotiation between the two parties affected. The privately negotiated alternatives seem more efficient than the mayor's rulings because the production of both candy and health care continues. Note that in these alternatives, both parties alter their behavior. For example, the doctors take a break, and the candy factory limits loud noise to the afternoon. Thus, the parties find a solution in which the polluter does not make all the adjustments, as would be the case if the mayor adopted a "no loud noise" rule.

property rights: rights over the use, sale, and proceeds from a good or resource. (Ch. 1)

The Importance of Assigning Property Rights. For a negotiation like this to work, however, it is essential that property rights be well defined. *Property rights* determine who has the right to pollute or infringe on whom. Who, for example, is being infringed on in the case of the noise pollution? Does the candy factory have the right to use loud machinery, or does the hospital have the right to peace and quiet? The mayor's ruling could establish who has the property right, but more likely the case would be taken to a court and the court would decide. After many such cases, precedent would establish who has the property rights in future cases.

Coase theorem: the idea that private negotiations between people will lead to an efficient resolution of externalities regardless of who has the property rights as long as the property rights are defined.

The property rights will determine who actually pays for the adjustment that remedies the externality. If the candy factory has the right, then the workers can demand some compensation (perhaps free health-care services) from the hospital for limiting their noise in the afternoon. If the hospital has the right, then perhaps the doctors can get compensated with free candy during the break. The **Coase theorem** states that no matter who is assigned the property rights, the negotiations will lead to an efficient outcome as described in the candy/health-care example. The assignment of the property rights determines who makes the compensation.

transaction costs: the costs of buying or selling in a market, including search, bargaining, and writing contracts.

Transaction Costs. Even if property rights are well defined, for a private agreement like this to occur, transaction costs associated with the agreement must be small compared to the costs of the externality itself. **Transaction costs** are the time and effort needed to reach an agreement. As Coase put it, "in order to carry out a market transaction, it is necessary to discover who it is that one wishes to deal with, to inform people that one wishes to deal and on what terms, to conduct negotiations leading up to a bargain, to draw up the contract, to undertake the inspection needed to make sure that the terms of the contract are being observed, and so on. These operations are often extremely costly."[2] Real-world negotiations are clearly time-consuming, requiring skilled and expensive lawyers in many cases. If these negotiation costs are large, then the private parties may not be able to reach an agreement. If the negotiation in the health-care/candy example took many years and had to be repeated many times, then it might be better to adopt a simple "no loud noise" rule.

■ **The Free-Rider Problem Again.** Free-rider problems can also prevent a private agreement from taking place. For example, a free-rider problem might occur if the hospital was very large, say, 400 doctors. Suppose that the candy workers have the right to noise pollute, so that they require a payment in the form of health care. The hospital would need contributions from the doctors to provide the care. Thus, if each doctor worked in the hospital an extra day a year, this might be sufficient.

However, any one of the 400 doctors could refuse to work the extra day. Some of the doctors could say that they have other job opportunities where they do not have to work an extra day. In other words, doctors who did not pay could free-ride: work at the hospital and still benefit from the agreement. Because of this free-rider problem, the hospital might find it hard to provide health care to the candy workers, and a private settlement might be impossible.

Thus, in the case where the transaction costs are high or free-rider problems exist, a private remedy may not be feasible. Then the role of government comes into play, much as it did in the case of public goods, where the free-rider problem was significant. Again as Coase put it, "Instead of instituting a legal system of rights which can be modified by transactions on the market, the government may impose regulations which state what people must or must not do and which have to be obeyed."[3]

Command and Control Remedies

When private remedies for externalities are either too costly or not feasible because of free-rider problems, there is a role for government. One form of government intervention to solve the problem of externalities is the placement of restrictions or regulations on individuals or firms, often referred to as **command and control.** Such restrictions could make it illegal to pollute more than a certain amount. Firms that polluted more than that amount could then be fined. For example, in the United States, the corporate average fleet efficiency (CAFE) standards require that the fleet of cars produced by automobile manufacturers each year achieve a stated number of miles per gallon on the average. Another example is a government requirement that electrical utilities put "scrubbers" in their smokestacks to remove certain pollutants from the smoke they emit. In this case, the government regulates the technology that the firms use. Reducing pollution by regulating what firms or individuals produce is a classic example of command and control. Through commands, the government controls what the private sector does. In principle, the externalities are made internal to the firm by requiring that the firm act as if it took the external costs into account.

command and control: the regulations and restrictions that the government uses to correct market imperfections.

[2] Ronald Coase, "The Problem of Social Cost," *Journal of Law and Economics,* October 1960, Vol. 3, p. 15.

[3] Ronald Coase, "The Problem of Social Cost," *Journal of Law and Economics,* October 1960, Vol. 3, p. 17.

Command and control methods are used widely by agencies such as the Environmental Protection Agency (EPA), which has responsibility for federal environmental policy in the United States. There are many disadvantages to command and control in the environmental area, however, and economists have criticized such methods. The most significant disadvantage is that command and control does not allow firms to find other, cheaper ways to reduce pollution. Command and control ignores the incentives firms might have to discover cheaper technologies. For example, under command and control, electrical utilities have to install a scrubber even if there is a better, cheaper alternative. New machinery without a scrubber might be more efficient than installing a scrubber. Similarly, developing alternative fuels or simply raising the price of gasoline might be à cheaper way to reduce pollution than the CAFE standards.

Taxes and Subsidies

Because of these disadvantages, economists recommend alternatives to command and control techniques to reduce pollution or to reduce the inefficiencies due to other externalities. Taxes and subsidies are one such alternative. How do they work?

Goods that have negative externalities are taxed. When there are many drivers in a city, roads become congested, leading to traffic backups and delays. Each driver contributes to the congestion, imposing external costs on the other drivers. In 2003, a new tax, called a congestion charge, was imposed on vehicles that drive in central London during the day. The idea was to reduce congestion in central London by making it more expensive to drive there. Cameras mounted above the roads check vehicle registration numbers to make sure that the tax has been paid. This tax internalizes the externality by making drivers pay for the external congestion costs through the tax. If the demand for days of driving in central London is downward-sloping, the tax will reduce driving in central London, reduce the external costs imposed on drivers, and create government revenue.

If a good has a positive externality, too little of it is consumed in a competitive economy. A subsidy can be used to increase consumption of the good. In 2000, the state of Arizona gave a $20,000 subsidy to people who bought an SUV that runs on both gasoline and propane. The idea was to give an incentive for people to buy vehicles that pollute less. The pollution reduction would then benefit all Arizona residents. This subsidy internalizes the externality by making the consumer feel the extra benefits through the subsidy. Do you think Arizona residents responded to the change in the price of these SUVs? As you probably guessed, they did; so many Arizona residents took advantage of the subsidy that an emergency session of the state legislature was convened to curtail the program.

Big Brother Is Watching Your Car!
Closed circuit television cameras loom above traffic in central London to monitor the license plates of the estimated 250,000 cars entering the city each day. Drivers pay a "congestion charge" to drive in central London, making it more expensive to drive there, with the ultimate goal of reducing traffic backups.

Unlike command and control, taxes and subsidies allow firms or people to respond to price or cost changes. With changes in technology, a firm might find it could afford to pollute even less than would be allowed under a command and control guideline.

The way that taxes can be used to reduce pollution is illustrated graphically in Figure 15.3, which uses the same curves as Figure 15.1. Recall that the marginal social cost of production is greater than the marginal private cost, as viewed from the private firm, because the good pollutes. We know that taxes raise the marginal cost to the individual firm. They thereby shift up

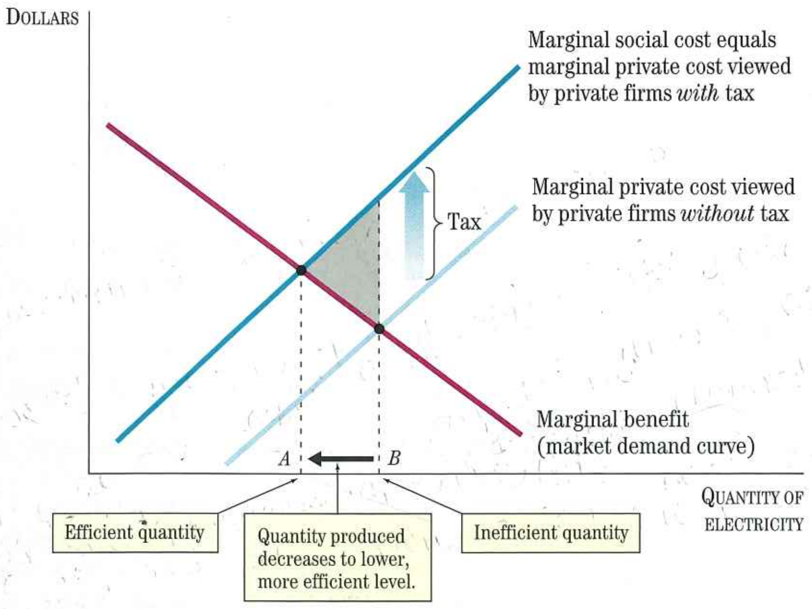


Figure 15.3
Using Taxes in the Case of a Negative Externality
A tax equal to the difference between the marginal private cost and the marginal social cost in Figure 15.1 shifts the supply curve up. This reduces the equilibrium quantity produced to the lower, more efficient level.

the market supply curve and lead to a market equilibrium with a smaller quantity produced. If the tax is chosen to exactly equal the difference between the marginal social cost and the marginal private cost, then the quantity produced will decline from the inefficient quantity shown at point *B* to the efficient quantity shown at point *A* in Figure 15.3.

There are many examples of taxes being used at least in part to reduce pollution. Gasoline taxes are widely viewed as being good for the environment because they reduce gasoline consumption, which pollutes the air. In the United States there is an average tax of 46 cents on each gallon of gasoline. The big advantage of taxes or subsidies compared with command and control is that the market is still being used. For example, if there is a shift in demand, the firm can adjust its technique of production as the price changes. But with command and control, adjustment must wait until the government changes its commands or controls.

In the case of positive externalities, subsidies rather than taxes can be used to increase production and bring marginal social benefits into line with marginal costs. For example, in Figure 15.4, which uses the same curve as Figure 15.2, a subsidy to encourage education, a good with a positive externality, is illustrated. In this case, a subsidy to students raises the marginal benefit of education (as perceived by them) up to the marginal social benefit. As a result, the quantity of education rises from the inefficient level (*C*) to the efficient level (*D*), as illustrated in Figure 15.4.

Externalities from Biodiversity

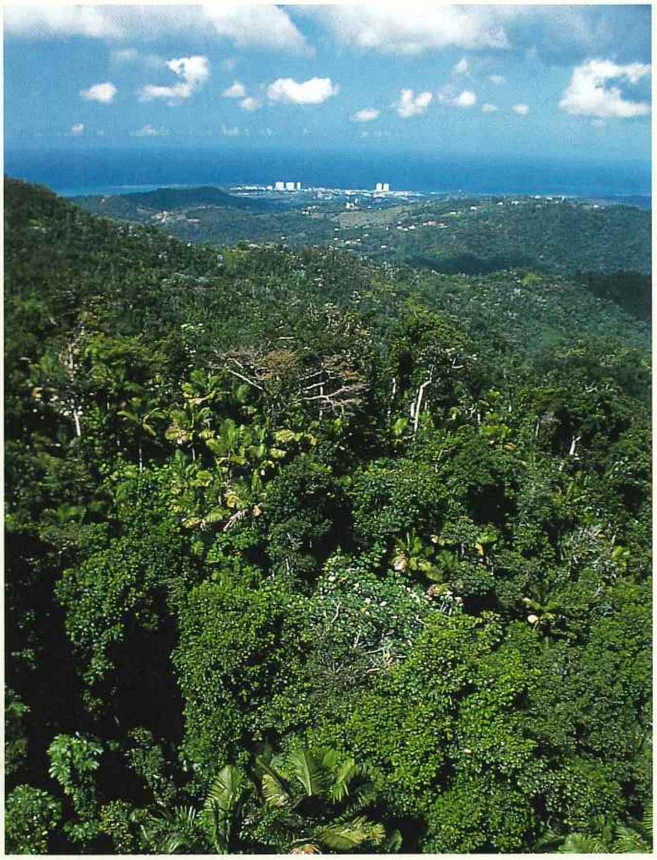
Biodiversity—the rich variety of plant and animal life in the world—has been recognized as important to maintaining the world's ecosystem. Any species may hold unique benefits for pharmaceutical and medical research. Many important pharmaceutical products throughout history—from aspirin to life-saving drugs—have been discovered in the natural environment and then modified or improved by researchers. Preserving biodiversity is important for future discoveries and applications.

One of the great sources of biodiversity is the rain forests of South America. However, these rain forests are being cut and burned to make room for farms.

Observe that there is an externality here. Those governments or individuals who own the rain forests may suffer little from cutting them down and losing the biodiversity. The benefit of the biodiversity is external to them, spread around the world and, indeed, to future generations, who must forgo the opportunity for better drugs that the variety of plant and animal life might bring. This externality is global, not restricted to any one country. Thus, resolving it is even more difficult than in the case of a single country with one government.

This externality is now being reduced by private remedies—negotiations between affected parties. In some cases, pharmaceutical companies have offered the owners of the rain forests an opportunity to share in some of the patent and copyright royalties from the discovery of new drugs. In exchange for not cutting down the forests, the owners of the forests can share in any royalties from drugs derived from plant and animal life from the forests. If such royalties are available, then by cutting down the forests the owners forgo the royalties; this raises the cost of cutting and burning and effectively internalizes the externality.

It is not yet clear whether the incentive will be great enough to slow the cutting and burning of the forests, or whether an international agreement among governments around the world is feasible. In fact, the difficulty of coordinating international government action in these cases may be the reason why interested private parties are looking for ways to resolve the externality themselves.



A benefit of this tropical rain forest may be a life-saving drug. Why is it an external benefit? How can it be internalized?

In addition to subsidizing education, the government subsidizes research, another good with a positive externality, by providing research grants to private firms and individuals. The National Science Foundation supports basic research, and the National Institutes of Health support medical research. In supporting research with a limited budget, it is important for the government to place more emphasis on research that has big externalities. Many view basic research as having larger positive externalities than applied research. The ideas in basic research, such as that on the structure of the atom, affect many parts of the economy. Applied research, such as that on a new lightweight metal for a bike, has more limited use, and the firm can prevent others from using it. Products developed through applied research can be

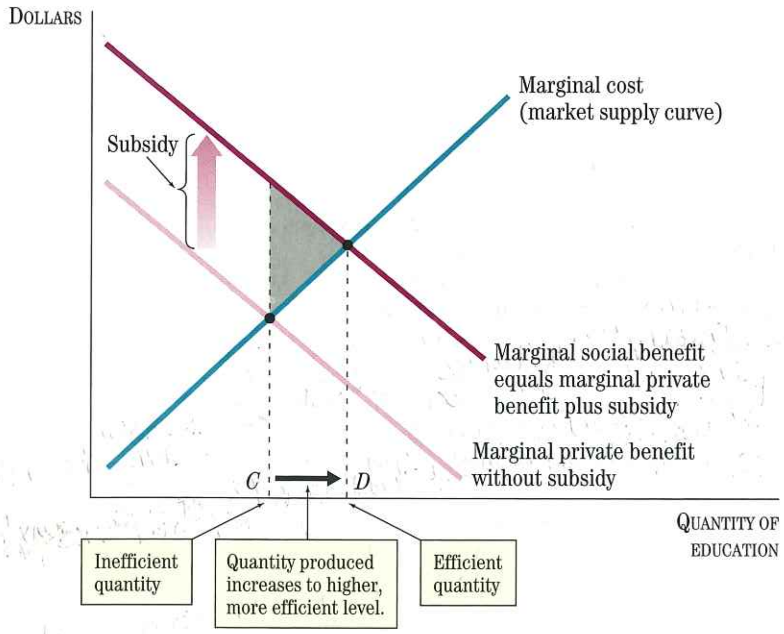


Figure 15.4
Using Subsidies in the Case of a Positive Externality
A subsidy equal to the difference between the marginal social benefit and the marginal private benefit of education or research shifts the demand curve up. This increases the equilibrium quantity produced to the higher, more efficient level and eliminates the deadweight loss due to the externality.

sold for profit. This suggests that more government funds should go toward basic research than toward applied research. In fact, the federal government in the United States does spend more to subsidize basic research than applied research.

Emission Taxes. A more direct way to use taxes to deal with pollution externalities is to tax the firm based on the amount of pollution emitted. For example, an electrical utility could be charged fees depending on how many particles of sulfur dioxide it emits, rather than on how much electricity it produces. Such charges are called **emission taxes.** They are much like taxes on the amount of the product sold, but they focus directly on the amount of pollution.

emission tax: a charge made to firms that pollute the environment based on the quantity of pollution they emit.

Emission taxes have an advantage over taxes on production in that the firm can use technology to change the amount of pollution associated with its production. Thus, rather than producing less electricity, the firm can reduce the amount of pollution associated with a given amount of electricity if it can find a cheaper way to do so. Emission taxes have an even greater advantage over command and control than a tax on the product has.

■ **Why Is Command and Control Used More Than Taxes?** There is one feature of command and control that many people like: The total amount of pollution can be better controlled than with a tax. This may explain why command and control is used more than taxes. Suppose, for example, that a tax is used to equate marginal social cost with marginal benefit, as in Figure 15.3. Now suppose there is a sudden reduction in the private cost of producing electricity. The private marginal cost curve will shift down and, with the tax unchanged, production (and pollution) will increase. A regulation that stipulates a certain quantity to produce would not have this problem. The total amount of pollution would be fixed. Fortunately, in recent years, a new idea in pollution control has emerged that has both this advantage of command and control and the flexibility of the market. This new idea is tradable permits.

Tradable Permits

tradable permit: a governmentally granted license to pollute that can be bought and sold.

Tradable permits use the market to help achieve the standards set by the government. Rather than force a firm to meet a certain standard, the government issues each firm a permit that allows the firm to emit a certain limited amount of pollutants into the atmosphere. Firms have an incentive to lower their emissions because they can sell the permit if they do not use it. Firms that can lower their emissions cheaply will choose to do so and benefit by selling their permits to other firms for which reducing the pollution is more costly. Tradable permits not only allow the market system to work, they give firms incentive to find the least costly form of pollution control.

■ **Control over Firms as a Group Rather Than Individual Firms.** Tradable permits are ideal in certain circumstances. For example, they work well in the case of acid rain, which falls over a wider area than that of the individual polluting firm. With acid rain, it is the total amount of pollution from all firms in the country or the region that matters most. To reduce the total amount of pollution, the government issues a number of permits specifying permissible levels of pollution. Once these permits are issued, those firms that can reduce the emissions in the most cost-efficient manner will sell their permits to other firms. They can raise profits by reducing pollution themselves and selling their permits to other firms with less efficient pollution-control methods. The total amount of pollution in the economy is equal to the total amount of permits issued and, therefore, is controlled perfectly.

Tradable permits are a new idea but are likely to be an increasingly common way to reduce pollution in the future. A tradable permit program called RECLAIM is being used in Los Angeles. Under the 1990 Clean Air Act, tradable permits can be used on a national basis.

Tradable permits could also work in global warming. The amount of global warming depends on the total amount of carbon dioxide emissions in the world's atmosphere. It does not matter whether a firm in Los Angeles or in Shanghai emits the carbon dioxide. Tradable permits could control the total amount of pollution. The permits would let firms or individuals decide on the most cost-effective way for them to reduce the total amount of pollution.

■ **Assigning and Defining Property Rights.** Tradable permits illustrate how important property rights are for resolving externalities. The role of government in this case is to create a market by defining certain rights to pollute and then allowing firms to buy and sell these rights. Once rights are assigned, the market can work and achieve efficiency.

Balancing the Costs and Benefits of Reducing Externalities

As with public goods, it is important to use a cost-benefit analysis when considering externalities. There are benefits to reducing pollution, but there are costs also. The costs of reducing pollution in the United States are about $120 billion per year, or about 2 percent of GDP. This percentage is expected to rise over time. These costs should be compared to the benefits associated with pollution control on a case-by-case basis.

For example, the Environmental Protection Agency introduced a new rule for stricter standards on the amount of sulfur allowed in diesel fuel in 2000. Environmentalists were concerned that diesel exhaust is causing accelerated cancer rates, but trucking companies that use diesel fuel were concerned about the effect of the stricter standards on the price of their fuel. The Environmental Protection Agency estimated that the new, stricter standards would prevent 8,300 deaths and 360,000 asthma attacks, while increasing the price of diesel fuel by four to five cents per gallon. The oil industry reported that implementation of the stricter standards would cost $8 billion.

The stricter standards save lives and reduce suffering for hundreds of thousands of people. These benefits should be compared to the cost of the stricter standards to determine environmental policy. Why was this rule change so hotly debated? You may have noticed that the citizens who avoid cancer and breathing problems benefit from the stricter standards, while the cost is borne by the oil producers and trucking companies.

■ **Environmental Policy Is Debated** The following two examples illustrate recent environmental policy debates, with a proposed change in an environmental rule, the reported benefits and costs of reducing the externality, and the actual outcome in each case. Imagine that you are in charge of determining U.S. environmental policy. Think about the benefits and the costs of the stricter regulation, and the tradeoff between costs and benefits. Would you decide that a new, stricter regulation should be in place to reduce the externality?

In the first example, the cost of reducing the externality was judged to be too high compared to the benefits. In the 2000 campaign for the presidency, George W. Bush supported the regulation of carbon dioxide emissions. Bush called carbon dioxide a pollutant that contributes to global warming. Environmental groups supported the regulation of carbon dioxide emissions and were enthusiastic about Bush's concern about global warming. In 2001, President Bush changed his mind about regulating carbon dioxide emissions. A study by the Energy Department showed that this regulation would result in nearly a quadrupling of the cost of producing electricity from coal, causing large price increases for both electricity and natural gas. President Bush stated that the impact on utility prices made the cost of regulating carbon dioxide emissions too high.

In this second example, the benefits from reducing the externality were judged to be worthwhile compared to the costs. Three days before leaving office, President Clinton lowered the standard for arsenic in drinking water by 80 percent. In March 2001, the Environmental Protection Agency rescinded the new, stricter standard and debated its merits. Scientific studies show that higher levels of arsenic in drinking water lead to higher risks of fatal cancer, heart disease, and diabetes. Arsenic is a by-product in mining, is used as a wood preservative for lumber, and occurs naturally in water in some areas. The cost of the stricter standard for arsenic in drinking water was estimated to be billions of dollars. In this case, the Bush administration

READING THE NEWS ABOUT... Environmental Economics

Economics has become a vital tool of environmentalists working to preserve the natural environment. Many have found that applying economic reasoning to issues that were once argued only on moral or ethical grounds can be a very effective way to persuade businesses and governments to consider environmental issues—and ultimately achieve the environmentalists' goals.

Green Groups See Potent Tool In Economics

By JESSICA E. VASCELLARO
Staff Reporter of THE WALL STREET JOURNAL
August 23, 2005

Many economists dream of getting high-paying jobs on Wall Street, at prestigious think tanks and universities or at powerful government agencies like the Federal Reserve.

But a growing number are choosing to use their skills not to track inflation or interest rates but to rescue rivers and trees. These are the "green economists," more formally known as environmental economists, who use economic arguments and systems to persuade companies to clean up pollution and to help conserve natural areas.

Working at dozens of advocacy groups and a myriad of state and federal environmental agencies, they are helping to formulate the intellectual framework behind approaches to protecting endangered species, reducing pollution and preventing climate change. They also are becoming a link between left-leaning advocacy groups and the public and private sectors.

"In the past, many advocacy groups interpreted economics as how to make a profit or maximize income," says Lawrence Goulder, a professor of environmental and resource economics at Stanford University in Stanford, Calif. "More economists are realizing that it offers a framework for resource allocation where resources are not only labor and capital but natural resources as well."

Environmental economists are on the payroll of government agencies (the Environmental Protection Agency had about 164 on staff in 2004, up 36% from 1995) and groups like the Wilderness Society, a Washington-based conservation group, which has four of them to work on projects such as assessing the economic impact of building off-road driving trails. Environmental Defense, also based in Washington, was one of the first environmental-advocacy groups to hire economists and now has about eight, who do such things as develop market incentives to address environmental problems like climate change and water shortages.

"There used to be this idea that we shouldn't have to monetize the environment because it is invaluable," says Caroline Alkire, who in 1991 joined the Wilderness Society, an advocacy group in Washington, D.C., as one of the group's first economists. "But if we are going to engage in debate on the Hill about drilling in the Arctic we need to be able to combat the financial arguments. We have to play that card or we are going to lose."

decided to impose the stricter standard, convinced that the health benefits were worth the cost.

Some people are concerned that a cost-benefit analysis will reduce spending on the environment too much. They argue that there is no tradeoff between costs and benefits. Environmental regulations can benefit rather than cost the economy, they argue, because requiring individuals to reduce pollution creates a demand for pollution-reducing devices and creates jobs in the pollution-reducing industry. But unless the pollution-reducing equipment is creating a benefit to society greater than

The field of environmental economics began to take form in the 1960s when academics started to apply the tools of economics to the nascent green movement. The discipline grew more popular throughout the 1980s when the Environmental Protection Agency adopted a system of tradable permits for phasing out leaded gasoline. It wasn't until the 1990 amendment to the Clean Air Act, however, that most environmentalists started to take economics seriously.

The amendment implemented a system of tradable allowances for acid rain, a program pushed by Environmental Defense. Under the law, plants that can reduce their emissions more cost-effectively may sell their allowances to more heavy polluters. Today, the program has exceeded its goal of reducing the amount of acid rain to half its 1980 level and is celebrated as evidence that markets can help achieve environmental goals.

Its success has convinced its former critics, who at the time contended that environmental regulation was a matter of ethics, not economics, and favored installing expensive acid rain removal technology in all power plants instead.

Greenpeace, the international environmental giant, was one of the leading opponents of the 1990 amendment. But Kert Davies, research director for Greenpeace USA, said its success and the lack of any significant action on climate policy throughout early 1990s brought the organization around to the concept. "We now believe that [tradable permits] are the most straightforward system of reducing emissions and creating the incentives necessary for massive reductions."

Organizations are also applying economic reasoning toward saving wildlife. In response to arguments that undeveloped land hurts economic growth, Defenders of Wildlife founded a conservation-economics program in 1999 and recently oversaw a study of how much tourists would be willing to pay to visit a red-wolf reservation and educational center in Columbia, N.C. The finding that the center's $2 million price tag would be paid by tourism revenue in five to 10 years is helping raise money for the center and being used by advocacy groups attempting to reintroduce the population in the area.

Environmentalists have also come to recognize that if they can couch their arguments in economic terms, not only governments but also corporations are more likely to listen. Since 2001, the San Francisco-based Rainforest Action Network has persuaded J.P. Morgan Chase & Co., Citigroup Inc. and Bank of America Corp. to account for the cost of pollution in their loan-underwriting processes and, in some cases, to avoid investing in industrial logging companies.

"Companies are looking for certainty and stability," says Michael Brune, executive director of the Rainforest Action Network. "They can do that by investing in sustainable energy, where they don't run the risk of lawsuits or federal regulation or the reputation of being associated with environmentally controversial projects."

the benefits of other goods, shifting more resources to pollution abatement will not be an efficient allocation of society's resources.

Many people argue that the surest way to reduce pollution around the world is to make sure the less-developed economies of the world increase their level of income. This will give them more resources to spend on pollution control; in fact, poor countries will not spend much on reducing pollution until the problems of poverty and hunger are reduced. Environmental degradation in Eastern Europe was severe when these economies were centrally planned. It is likely that the environment will improve when they have more resources to spend on it.

REVIEW

- There are four basic ways to improve the efficiency of markets in the case of externalities: private remedies, command and control, taxes and subsidies, and tradable permits. The latter three involve direct government intervention and are needed when high transaction costs or free-rider problems rule private remedies out.
- Taxes and subsidies, as well as tradable permits, endeavor to use the market to help internalize the externality.
- Tradable permits require that the government define and assign property rights. Once the rights are defined and assigned, the market can allocate the resources to reduce pollution efficiently.

Models of Government Behavior

The previous two sections have outlined what government should do to correct market failure due to public goods and externalities. Regardless of the reason market failure occurs, the outcome is similar: Production may be too little or too much, and producer surplus plus consumer surplus is not maximized. The result is deadweight loss, and the role of government is to change the level of production or employment so as to increase producer surplus plus consumer surplus. Using economics to explain the role of government in this way is considered a *normative* analysis of government policy. Normative economics is the study of what *should be* done. But there is another way to look at government policy. It falls into the area of *positive* rather than normative economics and looks at what governments *actually do* rather than what they should do.

One of the reasons for studying what governments actually do is that frequently the normative recommendations are not followed, or government performs its role poorly. **Government failure** occurs when the government fails to improve on the market or even makes things worse. Sometimes government fails, and sometimes it succeeds. One objective of positive analysis of government is to understand why there is success and failure in different situations.

government failure: the situation where the government fails to improve on the market or even makes things worse.

Public Choice Models

Government itself is run by people. Government behavior depends on the actions of voters, politicians, civil servants, and political appointees from judges to Cabinet officials. The work of government also depends on the large number of people who work in political campaigns, who are active in political parties, who lobby, and who participate in grassroots campaigns, from letter writing to e-mail messages to political protests. Government organizations exist at the state and local levels as well as at the federal level. What motivates the behavior of all these people?

The motivations of politicians and government workers are complex and varied. But the central idea of economics that people make purposeful choices with limited resources should apply to politics and government, as well as to consumers and firms. Many people enter politics for genuine patriotic reasons and are motivated by a desire to improve the well-being of people in their city, state, or country, or even the world. Their motivations may be deeper than watching out for their own best interests, narrowly defined. For example, Alexander Hamilton, the first chief economic spokesman for the United States as the first secretary of the treasury, worked hard to put the newly formed country on a firm economic foundation by having the federal government assume the debts of the states after the Revolutionary War.

But the desire to get elected, or to get votes on issues after being elected, is also part of the motivation of all politicians. Alexander Hamilton would not have done his job if he had not made one of the great political deals of all time, trading his vote on one issue for votes on another. In order to get the votes of the representatives from Virginia and Maryland for the federal government to assume the debts of the states, he agreed to vote to place the capital of the new country along the banks of the Potomac River between Maryland and Virginia, instead of selecting New York City.

public choice models: models of government behavior that assume that those in government take actions to maximize their own well-being, such as getting reelected.

Economic models of government behavior are called **public choice models.** They start from the premise that politicians are motivated by increasing their chances of getting themselves or the members of their party elected or reelected. And without explicit incentives to the contrary, government workers are presumed to be motivated by increasing their power or prestige, through increasing the size of their department or by getting promoted. By understanding this self-interest motivation, we can learn much about government, including the reasons for government failure and the reasons for government success.

Economic Policy Decisions Through Voting

Let us first examine how voting is used to make economic policy decisions in a political environment. We will use the assumption of public choice models: that getting elected is the primary motivation of politicians.

Single Issues with Unanimity. Let us start with the easiest case: There is only one economic policy decision to be made, and all the voters agree on what it should be. For example, suppose that the issue is spending on national defense, a public good where the government has a key role to play according to the normative economic analysis discussed earlier.

Suppose the specific issue is how much to spend on national defense now that the cold war is over. Some alternatives are shown in Table 15.2.

Table 15.2
Alternative Levels of National Defense Spending

National Defense as a Share of GDP	
1 percent	Japan's maximum
2 percent	U.S. in 1940
3 percent	U.S. in 2000
4 percent	Post–cold war
10 percent	U.S. in 1960
39 percent	U.S. in 1944

Suppose that everyone agrees that a level of national defense of around 4 percent of GDP in the United States is appropriate for the post–cold war period, in the absence of major world political changes such as the events of September 11. In reality, of course, opinions differ greatly about the appropriate level. But suppose that after looking at history or making international comparisons or listening to experts on defense and world politics, everyone agrees that 4 percent of GDP is the right amount to spend.

Under these circumstances, when there is only one issue on which all voters agree, voting will lead to the action that everyone prefers, that is, 4 percent, even if politicians are motivated by nothing other than getting elected. Suppose that one politician or political party runs for election on a plank of 39 percent defense spending and that the other argues in favor of 2 percent; clearly, the party with 2 percent will win because it is much closer to the people's views. But then the other politician or party will see the need to move toward the consensus and will run on a 5 percent spending platform; if the other party stays at 2 percent, then the higher-spending party will win. But clearly the other party will then try to get closer to 4 percent, and eventually 4 percent will be the winner.

This example shows that the political system yields the preferred outcome. Of course, after being elected, the politician might break the promise made during the campaign. But if such a change cannot be justified on the basis of a change in circumstances, that politician may have difficulty getting reelected.

The Median Voter Theorem. What if people have different views? Suppose there is no unanimity about a 4 percent share of GDP for defense. Instead, the country consists of people with many different opinions. Some want more than 4 percent; some want less than 4 percent. Suppose that about half of the people want more than 4 percent and half want less than 4 percent; in other words, 4 percent is the desire of the *median* voter.

If there is only one issue, there will be convergence of the positions of the politicians or the parties toward the median voter's belief. For example, if one party or politician calls for 7 percent spending and the other party calls for 4 percent, then the party calling for 4 percent will attract more voters. Clearly, more than half of the voters are closer to 4 percent than to 7 percent. The **median voter theorem** predicts that the politicians who run on what the median voter wants will be elected. The views of the people at the extremes will not matter at all.

median voter theorem: a theorem stating that the median or middle of political preferences will be reflected in government decisions.

Convergence of Positions in a Two-Party System. An interesting corollary to the median voter theorem is that political parties or politicians will gravitate toward the center of opinion—toward the median voter. For example, in the case of national defense, it makes no sense for any politician to run on a 39 percent recommendation. The parties will gravitate toward the median voter. This **convergence of positions** may explain the tendency for Democrats and Republicans to take similar positions on many issues.

convergence of positions: the concentration of the stances of political parties around the center of citizens' opinions.

Voting Paradoxes. When there are many different issues—defense, taxes, welfare, health-care reform—and people have different opinions and views about each issue, the outcome of voting becomes more complicated. Certain decision-making problems arise. The example of the **voting paradox** illustrates some of these problems.

voting paradox: a situation where voting patterns will not consistently reflect citizens' preferences because of multiple issues on which people vote.

Suppose three voters have different preferences on three different economic policy options—A, B, and C. Ali likes A best, B second best, and C the least; Betty likes B best, C second best, and A the least; and Camilla likes C best, A second best, and B the least. The three policy options could be three different levels of defense spending (high, medium, and low) or three different pollution control plans (emission taxes, tradable permits, and command and control). Table 15.3 shows the three voters and their different preferences on each option.

Table 15.3
Preferences That Generate a Voting Paradox

Ranking	Ali	Betty	Camilla
First	A	B	C
Second	B	C	A
Third	C	A	B

In voting on one option versus another, we get:
On A versus B: A wins 2 to 1
On B versus C: B wins 2 to 1
On A versus C: C wins 2 to 1

Paradox because A wins over B and B wins over C, yet C wins over A

Consider three different elections held at different points in time, each with one issue paired up against another. First, there is an election on A versus B, then on B versus C, and then on C versus A. The voting is by simple majority: The issue with the most votes wins. When the vote is on the alternatives A versus B, we see that A wins 2 to 1. That is, both Ali and Camilla like A better than B and vote for it, while only Betty likes B better than A and votes for B. When the vote is on B versus C, we see that B wins 2 to 1. Finally (this vote might be called for by a frustrated Camilla, who sees an opportunity), there is a vote on C versus A, and we see that now C wins 2 to 1. Although it looked like A was a winner over C—because A was preferred to B and B was preferred to C—we see that in the third vote, C is preferred to A; this is the paradox.

The voting paradox suggests that there might be instability in economic policies. Depending on how the votes were put together, the policy could shift from high defense to medium defense to low defense, or from one pollution control system to another, then to another, and then back again. Or taxes could be cut, then raised, and then raised again. All these changes could happen with nothing else in the world having changed. We could even imagine shifting between different economic systems involving different amounts of government intervention—from communism to capitalism to socialism to communism and back again!

This particular voting paradox has been known for two hundred years, but it is only relatively recently that we have come to know that the problem is not unique to this example. Kenneth Arrow showed that this type of paradox is common to any voting scheme. That no democratic voting scheme can avoid inefficiencies of the type described in the voting paradox is called the **Arrow impossibility theorem.**

Arrow impossibility theorem: a theorem that says that no democratic voting scheme can avoid a voting paradox.

The voting paradox suggests a certain inherent degree of instability in decisions made by government. Clearly, shifting between different tax systems frequently is a source of uncertainty and inefficiency. The voting paradox may be a reason for government failure in cases where the government takes on some activity such as correcting a market failure.

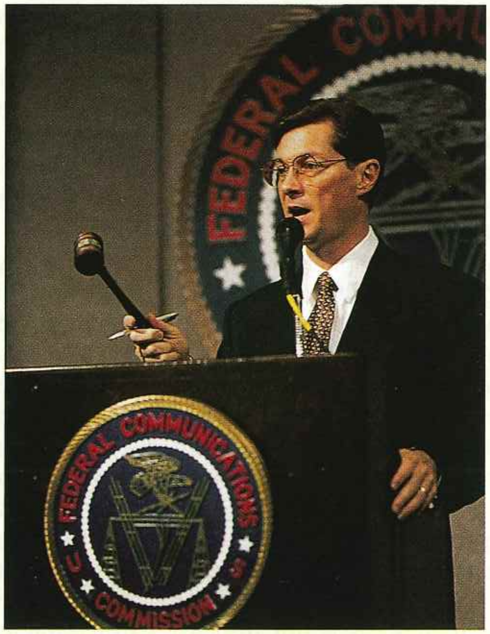
Special Interest Groups

The voting paradox is one reason for government failure. Special interest groups are another. It is not unusual for special interest groups to spend time and financial resources to influence legislation. They want policies that are good for them, even if the policies are not necessarily good for the country as a whole. For example, look at the farming industry, which has a great deal of government intervention. What is the explanation for the intervention? If you look back at the reasons for government intervention—income distribution, public goods, externalities—you will see that they do not apply to the farm sector. Food does not fit the definition of a public good, and many farmers who benefit from the intervention have higher incomes than other people in the society who do not benefit from such intervention. One can thus view the government regulation of agricultural markets as a form of government failure.

Concentrated Benefits and Diffuse Costs. One explanation for government failure in such situations is that special interest groups can have powerful effects on legislation that harms or benefits a small group of people a great deal but affects almost everyone else only a little. For example, the federal subsidy to the sugar growers in the United States costs taxpayers and consumers somewhere between $800 million and $2.5 billion per year, or about $3.20 to $10 per person per year. However, the gain from the subsidy amounts to about $136,000 per sugar

ECONOMICS IN ACTION

Advising the Government to Auction Off the Spectrum


The first FCC auction.

The U.S. government is responsible for distributing rights to use the radio-frequency spectrum in the United States. Each section, or band, of the spectrum is like a piece of property. Just as a farmer needs a piece of land to grow crops, a telecommunications firm needs a piece of the airwaves to send signals. And just as a piece of land has a price, so does a piece of the airwaves.

For many years, the U.S. government gave away the rights to use the spectrum in an arbitrary and inefficient manner. It was a classic example of elected officials and bureaucrats gaining influence or prestige by choosing who would get the rights. Economists had long recommended that government sell—auction off—the spectrum rather than give it away. They used models of government behavior to show why the traditional approach was inefficient and why it persisted. Finally, in 1993, Congress passed a bill giving the Federal Communications Commission (FCC) the authority to auction off the spectrum.

Auctioning off the airwaves is different from auctioning off art, however, because the value of a piece of the spectrum to firms depends on whether they also have adjacent parts of the spectrum—either adjacent geographically (like Florida and Georgia) or adjacent in frequency (with nearly the same megahertz number).

Because spectrum auctions were different, economists were called in to help design the spectrum auction. The auction design chosen by the FCC was a novel one. In most auctions, goods are auctioned off *sequentially*—first one piece of art, then the next, and so on. In contrast, following the advice of economists, bands of spectrum were auctioned off *simultaneously* by the FCC. In other words, firms could bid on several bands at the same time. It would be as if ten works of art were auctioned off at the same time, with buyers able to offer different bids on each piece of art. Thus, if the bids on one piece were too high, a buyer could change the bid on another piece before the final sale was made. This simultaneous procedure dealt with the distinct characteristics of the spectrum, namely, that many buyers wanted adjacent bands rather than a single band.

Because such a simultaneous auction had never taken place before, economic experiments were used to try it out. For example, Charles Plott of the California Institute of Technology conducted experiments on simultaneous auction proposals made by Paul Milgrom and Robert Wilson of Stanford University. Partly because the proposal worked well in the experiments, the FCC decided to use this approach. The auction process has been heralded as a great success, and the FCC is expected to continue auctioning the use of the spectrum. The Congressional Budget Office estimates that the revenue from these auctions in 2003–2005 will be around $24 billion.

grower. Thus, the small cost is hardly enough to lead each consumer to spend time fighting Congress. However, the payments are certainly worth the sugar growers' effort to travel to Washington and to contribute to some political campaigns. When the costs are spread over millions of users and the benefits are concentrated on only a few, it is hard to eliminate government programs. Those who benefit have much more incentive to lobby and work hard for or against certain candidates. Thus, the process of obtaining funds for election or getting support from the powerful interest groups can have large effects on policy.

■ **Wasteful Lobbying.** There is another economic harm from special interest lobbying. It is the waste of time and resources that the lobbying entails. Lobbyists are usually highly talented and skilled people, and millions of dollars in resources are spent on lobbying for legislation or other government actions.

In many less-developed countries—where special interest lobbying is more prevalent than in the United States—such activity consumes a significant amount of scarce resources.

Incentive Problems in Government

In any large government, many of the services are provided by civil servants rather than politicians and political appointees. In fact, it was to avoid the scandals of the spoils system—in which politicians would reward those who helped in a political campaign with jobs—that the civil service system set rules to protect against firing and established examinations and other criteria for qualifying workers for jobs.

But what motivates government managers and workers? Profit maximization as in the case of business firms is not a factor. Perhaps increasing the size of the agency or the department of government is the goal of managers. But simply increasing the size of an agency is not likely to result in an efficient delivery of services. Profit motives and competition with other firms give private firms an incentive to keep costs down and look for innovative production techniques and new products. But these incentives do not automatically arise in government. For this reason, it is likely that a government service, whether a public good or a regulation, will not be provided as efficiently as a good provided by the private sector. This is another possible reason for government failure.

Better Government Through Market-Based Incentives

In recent years, there has been an effort to use incentives to improve the efficiency of government. Many of these ideas were summarized in a popular 1992 book, *Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector*, by David Osborne and Ted Gaebler, which lent its name to the "reinventing government" movement of the 1990s.

Admitting that "cynicism about government runs deep within the American soul" and that "our government is in deep trouble today," the authors give hundreds of examples of how the "entrepreneurial spirit" can be used to make police services, sanitation, and schools more efficient. In many cases, efficiency can be improved by having government workers rewarded for providing high-quality service with higher pay or other benefits. In other words, marketlike incentives would be used to encourage greater government efficiency.

A big part of improving government can come through providing competition. Vouchers—including food stamps, housing vouchers, college tuition grants, elementary school grants—have been suggested by economists as a way to add competition and improve government efficiency. For example, Osborne and Gaebler contrast two different systems of government support for World War II veterans: (1) the GI bill, which gave veterans vouchers to go to any college, private or public, and (2) the Veterans Administration hospitals, where the government itself provides medical service. They conclude that the first system worked much better, and by analogy should be used in other cases where vouchers or government-produced services are the choices.

REVIEW

- Public choice models of government behavior assume that politicians and government workers endeavor to improve their own well-being, much as models of firms and consumer behavior assume firms and consumers do.
- In cases where there is consensus among voters, voting will bring about the consensus government policy. When there is no consensus, the median voter theorem shows that the center of opinion is what matters for decisions. However, the voting paradox points out that in more complex decisions with many options, the decisions can be unstable, leading to government failure.
- Other causes for government failure include special interest groups and poor incentives in government.
- Economic models of government behavior suggest ways to reduce the likelihood of government failure and increase government efficiency.
- Incentives and competition have been suggested as ways to improve the operation of government.

Conclusion

In this chapter, we have explored market failure due to public goods and to externalities. A competitive market provides too little in the way of public goods such as national defense and too little in the way of goods for which there are positive externalities, such as education and research. A competitive market results in too much production of goods for which there are negative externalities, such as goods that pollute the environment.

Most of the remedies for market failure involve the action of government. The provision of public goods by the government should require a careful cost-benefit analysis to make sure that the benefits are greater than the cost of producing a public good. The opportunities for private parties to work out externalities may be limited by transaction costs and free-rider problems. But there are ways in which the market system can aid the government, as in the case of tradable permits. In these cases, the main role of the government is to define and assign property rights.

It is very important, however, to develop models of government behavior and to recognize the possibility of government failure. In reality, political considerations enter into the production of public goods. A member of Congress from one part of the country might push for a public works project in his or her local district in order to be reelected. Moreover, the externality argument emphasized in this chapter is frequently abused as a political device, providing justification for wasteful expenditures. Thus, finding ways to improve decision-making in government, such as through market-based incentives, is needed if government is to play its role in providing remedies for market failures.

KEY POINTS

1. Public goods are defined by two key characteristics, nonrivalry and nonexcludability. National defense and police services are examples of public goods.
2. The existence of public goods provides a role for government because competitive markets frequently have difficulty producing such goods in the efficient amount.

3. Cost-benefit analysis is a technique to decide how much of a public good should be produced. Measuring benefits and deciding how to discount the future are difficult in the case of public goods.
4. Externalities occur when the costs or benefits of a good spill over to other parts of the economy. They create another potential role for government.
5. Goods may have a positive externality or a negative externality.
6. Externalities can sometimes be internalized in the private sector without government. But in many cases, externalities require some government action.
7. Taxes and subsidies or tradable permits are preferred to command and control because the market can still transmit information and provide incentives.
8. Models of government behavior are based on the economic assumption that people try to improve their well-being. In the case of politicians, this usually means taking actions to improve their chances of being elected or reelected.
9. The median voter theorem and the voting paradox are some of the results of the analysis of voting; the latter suggests a reason for government failure.
10. Special interest groups and poor incentives are some of the other reasons for government failure.
11. Marketlike incentives and competition are ways suggested by economists to reduce government failure.

KEY TERMS

public good
nonrivalry
nonexcludability
free-rider problem
user fee
cost-benefit analysis
contingent valuation
externality
negative externality
positive externality
marginal private cost
marginal social cost
marginal private benefit
marginal social benefit
internalize
private remedy
Coase theorem
transaction cost
command and control
emission tax
tradable permit
public choice models
median voter theorem
convergence of positions
voting paradox
Arrow impossibility theorem

QUESTIONS FOR REVIEW

1. What types of goods are produced or supplied by the government at the federal, state, and local levels?
2. Why do nonexcludability and nonrivalry make production by private firms in a market difficult?
3. Why is it difficult to measure the benefits of public goods when deciding how much to produce?
4. What is the use of cost-benefit analysis in the case of public goods?
5. What is the difference between a positive externality and a negative externality?
6. Why are private remedies for externalities not always feasible?
7. What is the advantage of emission taxes over command and control?
8. How do subsidies for education remedy a market failure?
9. What is the difference between the median voter theorem and the voting paradox?
10. What are the similarities and differences between market failure and government failure?

PROBLEMS

1. The following table shows the marginal benefit per year (in dollars) to all the households in a small community from the hiring of additional firefighters. The table also shows the marginal cost per year (in dollars) of hiring additional firefighters per year.

Number of Firefighters	Marginal Benefit	Marginal Cost
1	1,000,000	34,000
2	500,000	35,000
3	300,000	36,000
4	100,000	37,000
5	70,000	38,000
6	50,000	39,000
7	40,000	40,000
8	30,000	41,000
9	20,000	42,000
10	10,000	44,000

a. Is the service provided by the additional firefighters a public good?
b. Why might the marginal benefit from an additional firefighter decline with the number of firefighters?

c. Plot the marginal benefit and the marginal cost in a graph, with the number of firefighters on the horizontal axis.

d. What is the optimal amount of this public good (in terms of the number of firefighters)? Illustrate your answer on the graph in part (c).

e. Is this marginal benefit schedule the same as the town's demand curve for firefighters?

2. Suppose that there are only three households in the town in problem 1 and that each one of them has the marginal benefit (in dollars) from additional firefighters described in the following table:

Number of Firefighters	Household A	Household B	Household C
1	500,000	300,000	200,000
2	300,000	100,000	100,000
3	200,000	50,000	50,000
4	50,000	30,000	20,000
5	36,000	20,000	14,000
6	25,000	15,000	10,000
7	20,000	14,000	6,000
8	15,000	13,000	2,000
9	10,000	9,000	1,000
10	5,000	4,500	500

a. Add up the marginal benefits of the three households for each number of firefighters. Check that your addition gives the same marginal benefit for all the households in the town as given in problem 1.

b. Plot each of the three marginal benefit schedules and the marginal benefit schedule for the whole town on the same graph, with the number of firefighters on the horizontal axis. (You will need a big vertical scale.) What is the relationship between the three household curves and the curve for the whole town?

3. Suppose there is a neighborhood crime watch in which people volunteer to patrol the street where you live. If you do not participate in the patrol, but your neighborhood is safer because of the crime watch, are you a free rider? Why? What can your neighbors do to eliminate the free-rider problem?

4. Public education is not a public good, but it has external effects. Explain.

5. Group projects—for example, when students are assigned to work together on the same term paper—can lead to a free-rider problem. Why? What are some methods that teachers use to alleviate this free-rider problem?

6. Suppose that people value the continued existence of dolphins in the Pacific Ocean, but that tuna-fishing fleets kill large numbers of these mammals. Draw a graph showing the externality. Describe two alternative approaches to remedy the externality.

7. Alice's neighbors across the street want her to help them tend the flower garden in front of their house. Why is the flower garden an externality to Alice? What does this mean about the quantity of flowers that will be planted in the neighborhood? If Alice is planning to sell her house soon, will she be more or less likely to help her neighbors? Why?

8. Property rights over the world's oceans are not well defined. Recently, experts have noted that stocks of fish are declining as the seas' resources are overused.

a. Explain, in economic terms, why this might have happened.

b. Commercial fishing firms all over the world are complaining about the decline in their industry. The response of many governments has been to subsidize the fleets in their countries. Explain why this is an example of government failure.

9. In an attempt to set user fees, a state surveyed a sample of households about willingness to pay for camping in a particular state forest. The following table shows the results of this survey:

Price per Visit (in dollars)	Number of Visits (households per year)
50	0
45	1
40	2
35	4
30	6
25	10
20	20
15	100
10	160
8	200
7	400
6	600
5	1,000
4	2,000
3	6,000
2	8,000
1	10,000
0	20,000

a. What would the consumer surplus be each year if the price were $10 per visit?

b. To improve the camping facilities in this forest, the state is considering charging a fee of $5 per visit. Based on a cost-benefit analysis, what should the state take into consideration before improving the facilities?

10. List one specific example of a market failure and one of a government failure. What does the government do in the case of this market failure? Is the government successful? How might the market be used to reduce this

government failure? Is the government trying to correct this problem? What advice would you give to the government?

11. Cite one issue on which Republicans and Democrats have a convergence of positions and one on which the parties' positions are quite different. Why is there a difference between the issues you have selected?

12. Use the median voter theorem to explain why the admission of a group of extreme right-wing students to a college in place of a more moderate group of right-wing students will not affect the election of the class president.

13. Use the set of preferences for Ali, Betty, and Camilla, shown in the table in the next column, to show that the paradox of voting does not always occur.

	Ali	Betty	Camilla
First	A	B	C
Second	B	A	A
Third	C	C	B

How does this example differ from the one in Table 15.3?

14. After September 11, the media paid closer attention to the potential use of shipped cargo containers to smuggle weapons and terrorists into the United States.
    a. Discuss why there may be a negative externality associated with the use of shipping containers.
    b. Graphically show the market for ship transportation of cargo. Make sure you include all the relevant aspects of this market, given your answer to part (a). Explain verbally as necessary.
    c. Graphically show any deadweight loss that occurs in this market.
    d. Discuss potential ways of reducing or eliminating deadweight loss in this market and the advantages and disadvantages of these remedies, including their impact on other markets.