CHAPTER 8

Supply and Demand:
How Markets Work

bn Battuta, the great 14th-century Arab geographer, reported that long-distance trade occurred as follows along the Volga River in what is modern-day Russia:

> Each traveler . . . leaves the goods he has brought . . . and [the travelers] retire to their camping ground. Next day they go back to . . . their goods and find opposite them skins of sable, miniver, and ermine. If the merchant is satisfied with the exchange he takes them, but if not he leaves them. The inhabitants then add more skins, but sometimes they take away their goods and leave the merchant's. This is their method of commerce. Those who go there do not know whom they are trading with or whether they be *jinn* [phantoms] or men, for they never see anyone.[1]

Herodotus, the ancient Greek historian, describes similar exchanges—called *silent trade*—between the Carthaginians and the people of Libya in the fifth century BC. After having left their goods, Herodotus reports, the Carthaginians withdrew, and the Libyans "put some gold on the ground for the goods, and then pull back away from the goods. At that point the Carthaginians . . . have a look, and if they think there is enough gold to pay for the cargo they take it and leave." Herodotus describes how the process continued until an acceptable price was hit upon, remarking with surprise that "neither side cheats the other . . . [The Carthaginians] do not touch the gold until it is equal in value to the cargo, and the natives do not touch the goods until the Carthaginians have taken the gold."[2]

---

[1] Ibn Battuta, *Travels in Asia and Africa: 1325–1354* (London: Routledge and Kegan Paul, 1929), p. 151.

[2] Herodotus, *The Histories* (New York: Oxford University Press, 1998), pp. 300–301.

Silent trade is but one of the many ways that people have devised to engage in the process of exchange. Transfers of goods among strangers can range from gifts at one extreme, through mutually advantageous exchanges, to what might be called plunder at the other extreme. The potential gains from trade are often greater the more distant geographically or socially the parties to the exchange are. The fact that the parties to a silent trade did not meet face to face helped to reduce the chances of outbreaks of violence among the often heavily armed traders.

Other kinds of trade are anything but silent. The trading floor of one of the world's stock markets is a din of offers and bids, and a fruit market in modern Nigeria resounds with the almost musical call of market women advertising their wares and the hum of haggling over prices. Other modern markets are as silent and anonymous as the silent trade. When you buy a shirt or a book online, the only sound you hear is the mouse click when you hit the shopping cart icon on your screen. Similarly, you can buy an entire basket of groceries at a supermarket without saying a word, and, in contrast to the Nigerian fruit market, the only need for verbal communication occurs when the checker asks if you want plastic or paper bags for your goods! If you buy an item online through an auction market (such as e-Bay), you will experience an entirely different kind of market: there you will compete with others in posting a price for a good, the sale going to the highest bidder.

Goods and services are exchanged in many different ways. Families exchange gifts at holiday times, individuals work at jobs in return for money, and one member of a couple cares for the kids while the other works for the wages that pay the rent. In each case, who gets what in return for what is determined in a particular way, sometimes by custom, sometimes by law, and sometimes by the competitive forces of supply and demand. The main idea of this chapter is that *competitive markets—an important form of exchange—can be analyzed using the concepts of supply and demand.* This idea is expressed in two key points:

1. *Competitive markets* consist of many potential buyers and sellers, each acting independently, with no one participant having enough power to dictate terms to any other.

2. The key concepts in understanding how competitive markets work are *supply curves, demand curves,* and *market-clearing.*

THE NATURE OF MARKETS

A *market exchange* occurs when the owner of a good or service sells it to someone else. Selling it usually means that it is exchanged for money: the seller transfers ownership of the item to the buyer and receives money in return. The other side of the transaction is that the buyer pays money to acquire ownership of the good or service. If money is not involved in a transaction—if one product is traded directly for another—this is called *barter.*

> A **market exchange** is a transfer of title to a piece of property (a good or a service) to another party in return for some form of payment at mutually acceptable terms.

A market exchange is thus a transfer of title to a piece of property (a good or a service) to another party in return for some form of payment on mutually accepted terms. To say that the terms of a market exchange are mutually accepted means that under the circumstances, both parties

would rather make the exchange on the terms that are being offered than make no exchange at all.

People always have the right to refuse a market exchange. Therefore, by definition, all market exchanges that take place can be said to be *voluntary* in nature. But sometimes the right to walk away from an exchange does not mean much. For example, a severe illness in your family may force you to sell your home to pay the medical bills. It is not that someone is holding a gun to your head and telling you to sell your home, but sometimes circumstances can force people into exchanges they would otherwise avoid. (Recall Marlon Brando in "The Godfather" saying to one of his henchmen: "Make him an offer he can't refuse.")

The term *market* refers to the buying and selling activities of all those who want to trade (buy or sell) a particular good or service. Market activities are sometimes but not

> A **market** refers to all the buying and selling activities of those persons wishing to trade a good or service; a market consists of suppliers wanting to sell and demanders wanting to buy.

always concentrated in one location. Examples of markets with specific locations are the New York Stock Exchange and the Tokyo fish market. In these cases you can see the buying and selling of stocks or fish in one location. For other markets, however, there is no single specific place where you can "see" the market. For example, the Chicago labor market includes all the potential buyers and sellers of labor time who are meeting and coming to terms (or not coming to terms) anywhere in the Chicago area. A market, then, is not a place but rather a set of buying and selling activities.

Markets work to determine two basic economic outcomes: the *price* at which a good or service is exchanged and the *quantity* of it that will be bought and sold. These two outcomes affect many other aspects of society. The labor market, for example, determines not only wages (and hence living standards) but also the amount of employment (and hence also the number unemployed).

Each market has two types of participants: *demanders,* or those wishing to buy the good or service, and *suppliers,* or those wishing to sell the good or service. A market may comprise, say, two potential demanders facing three potential suppliers. This might occur in a local real estate market. Or the market might have a small number of suppliers and millions of demanders, as in the computer industry. Some markets have thousands of suppliers and just a few demanders, as in the labor market of a town with just a few large employers.

In this chapter we focus on markets with large numbers of potential demanders and suppliers. Following in the tradition of Adam Smith, such markets are termed *competitive markets* since the rivalry of the different participants—each one competing to make an ad-

> **Competitive markets** are those with many actual or potential demanders and suppliers.

vantageous purchase or sale—greatly affects the actions of all the others. Many markets in the United States and throughout the world are not competitive in this sense. We explain the workings of markets with smaller numbers of competitors in Chapter 11.

The most important consequence of having large numbers of participants in competitive markets is that no one of them is powerful enough to influence the price at which goods will sell. If there were just one seller, for example, a large corporation, it could gain a higher price for its product by making less of it available. But this strategy is ruled out in competitive markets.

SUPPLY AND DEMAND

We can understand how markets work by looking at the interaction of demanders and suppliers. We do this with the help of demand curves and supply curves.

A *demand curve* is a graphical representation of the buyers' side of the market. It shows how much of a particular commodity the demanders of this product will want to purchase at each possible price, given their taste for the product and the amount of money they have at their disposal. Each point on the curve represents a particular combination of a price (represented on the vertical axis) and the corresponding quantity demanded (measured on the horizontal axis).

> A **demand curve** indicates, for each possible price, how much of the good or service demanders are willing and able to buy.

Consider, as an example, the market for a certain item, say, beer in Iowa City, Iowa. (In the remainder of this chapter we assume that the beer referred to is all of the same type and quality; in the jargon of economics, we are assuming that beer is a *homogeneous* commodity.) Imagine that we asked every person in Iowa City (and all those who might travel to Iowa City to buy beer), "How many bottles of beer would you buy today if the price were $2 per bottle?" We would then add up all the answers. If the total came to 1,040 bottles, we would have one point on the demand curve: at a price of $2, buyers will demand 1,040 bottles on this day.

We might then repeat the survey, asking buyers how many bottles they would buy, first, if the price were $1 and, second, if it were $0.50 per bottle. Suppose we obtained answers of 2,000 bottles at a price of $1 and 3,760 at $0.50. We would then have two more points on the demand curve for beer in Iowa City on this day.

In Figure 8.1 the demand curve DD shows the various quantities of beer that buyers in Iowa City will demand on a certain day at all the possible prices, including the prices of $2, $1, and $0.50 for which we obtained answers in our survey. It is important to remember that both the demand curve and the supply curve present answers to *hypothetical* questions. In the case of the demand curve, the question is, "If the price were to be _____, what quantity would you buy today?" As we will explain shortly, most of the combinations of price and quantity on the demand curve and the supply curve will not actually be chosen.

Demand curves are almost always thought of as sloping downward to the right, or having a negative slope, as does DD in Figure 8.1. The economic reason for this is that, in general, the lower the price, the more of the good buyers will want to buy. In our beer example (Figure 8.1), if the price is high, say, $2 per bottle, then consumers will want to buy relatively few bottles per day. If the price is low, say, $0.50 per bottle, they will buy a much larger quantity each day.

Note that the demand curve does not represent what buyers *need*. It reflects only what they *want* and are *able to purchase,* given the *price* and their *incomes.*) Adam Smith defined what is reflected in a market as "effectual demand" (saying that a poor man's wish for a carriage drawn by six horses would not be an effectual demand), and modern economists have followed in his footsteps, defining "demand" simply as a want backed up by money.[3] We cannot tell from DD whether the buyers of beer are desperately thirsty after performing

[3] Adam Smith, *Wealth of Nations,* Book I, Chapter VII, eighth paragraph.
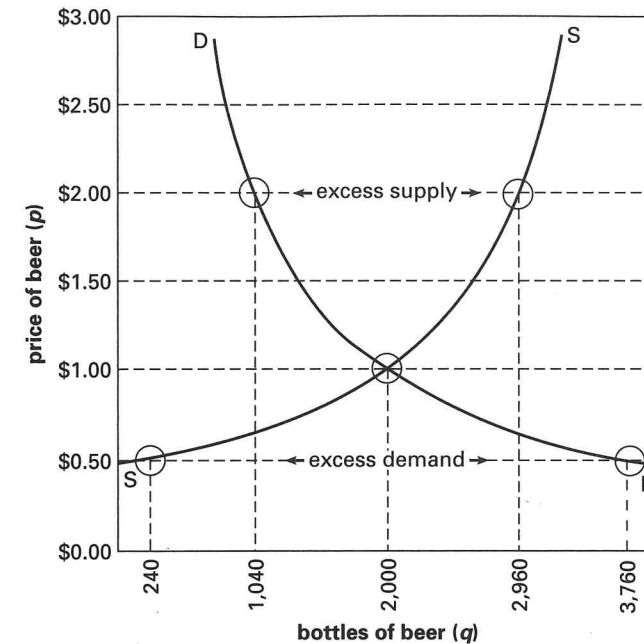

FIGURE 8.1   The demand curve and the supply curve.   The demand curve DD provides the following information. If the price of beer per bottle is p dollars, the amount demanded by consumers will be q bottles a day. For example, if the price is $2 a bottle, consumers will buy 1,040 bottles a day. The supply curve SS provides similar information about what producers will offer on the market. If the price is p dollars per bottle, suppliers will offer q bottles per day on the market. For example, if the price is $2 a bottle, beer suppliers will want to sell 2,960 bottles a day. Normally, demand curves slope downward (to the right) and supply curves slope upward.

arduous labor or whether they have already had more beers than they should have had. Indeed, if there are some people in Iowa City who desperately want beer but have no money to buy it, their wants will not be expressed in DD at all, since the demand curve expresses only what people are willing and able to buy. All the demand curve tells us is how much beer consumers will buy at any given price.

How much beer people want to buy will depend on many considerations besides the price. As we have noted, it will depend on the buyers' incomes. If everybody suddenly got a pay raise, people might want to buy more of many things, including beer. Another factor is the number of potential buyers. If new people came to town, that would tend to increase the demand for beer. More importantly, demand will change with the strength of people's desire—or, as economists put it, their "preference"—for a product. Thus, a consumer in Iowa City may want to buy more beer if she recently watched an effective beer ad on TV. A final variable is the price of a competing product. For example, people might buy less beer if the price of wine suddenly dropped. Such products are termed *substitutes.*

A demand curve, then, expresses how much the buyers are willing and able to buy at the various possible prices, *assuming* that nothing else affecting their demand changes. If

nothing else changes, the demand curve allows us to say how a change in the price will affect the quantity demanded. For instance, in Figure 8.1 we can see that if the price drops from $2 to $1 per bottle, *and if nothing else changes,* the quantity demanded will rise from 1,040 to 2,000 bottles a day.

The *supply curve,* by contrast, represents the sellers' side of the market. It depicts the suppliers' willingness to sell beer at different prices, and this willingness to supply beer will depend to a large extent on the costs incurred by beer-producing firms. In Figure 8.1 the supply curve SS shows, for a particular day, what quantity of beer sellers will supply to the market at various prices. For instance, if the price were to be $2 on some day, suppliers would try to sell 2,960 bottles that day; at a price of $0.50, suppliers would try to sell 240 bottles; and so on.

> A **supply curve** indicates, for each possible price, how much of the good or service suppliers wish to sell.

Supply curves are almost always thought of as sloping upward to the right (having a positive slope), as does the supply curve SS in Figure 8.1. When prices are high, suppliers will want to sell a lot of beer compared to when prices are low. At higher prices it will pay to put on extra shifts of workers at the brewery. The high price may also attract new suppliers from nearby cities where sellers are not able to get such a high price. When the price is low, on the other hand, some suppliers in Iowa City may try to find other cities in which to sell their product; some of them might even stop producing beer.

Other factors besides price influence how much of a product suppliers will want to sell at each price. The cost of producing beer compared with the rewards available from producing and selling other items will affect how much beer the suppliers will want to sell at each price. For example, if a labor-saving technical change occurs in the brewery industry, the cost of producing beer will fall, and the profits to be made in brewing and selling beer will rise. As a result more firms will be attracted into the industry and the amount supplied will increase. Similarly, if the grain used in making beer becomes more abundant, its price will drop, the costs of brewing will fall, and the quantity of beer supplied at each price will increase.

A supply curve, then, represents the quantities of a commodity that sellers will supply to the market at various possible prices, assuming that everything else affecting its supply remains unchanged. If nothing but the price changes, the supply curve tells us how the quantity of the good supplied will change when the price changes. In Figure 8.1, for example, if the price of beer rises from $1 to $2 per bottle on a certain day, suppliers will want to increase the amount they offer in the market that day from 2,000 to 2,960 bottles.

A complete list of factors determining the position of the supply curve would include

- the technologies available for producing the good;
- the costs of the various inputs and how these may vary with the amounts of the inputs used;
- the cost of obtaining the necessary permission to reproduce something from a copyright or patent holder (if the firm's activity requires it), which means the right to use someone's idea, invention, or other creative product (such as a musical composition or work of art) that is patented or copyrighted; and
- the number of firms producing the good, including those that may enter or leave the industry if the price rises or falls enough to attract new firms or cause existing firms to cease producing the good.

You may want to test your understanding of the supply curve by asking yourself how changes in any one of the above factors will shift the supply curve in Figure 8.1 either to the right or to the left. For example, what would happen to the supply curve if a new technology for producing beer were to be invented, one that would allow the same quantity of output to be produced with less labor?

All the influences that affect the position of the supply curve combine also to determine the *marginal cost* of a commodity's production. "Marginal cost" is defined as the *increase in the total cost incurred by a firm when it increases its total output of a commodity by one unit.*

> **Marginal cost** is the increase in the total cost incurred by a firm when it increases its total output of a commodity by one unit.

To grasp the significance of the concept of marginal cost consider the following step-by-step reasoning: Small movements up a supply curve show, at each point along the curve, how much the price of a good must be increased to induce some supplier of that good to produce and offer for sale in the market one additional unit of the good. Even when the market price of something goes up, however, we know that no firm will produce an additional unit of the good in question unless the higher price covers the cost of producing an additional unit of it. Since the cost to a firm of producing an additional unit of a good is, in fact, its marginal cost, we can say that a supply curve shows not only the amount of a good that will be supplied at each price but also what the marginal cost of the good is for at least one firm in the market.

Of course, what happens in a market is not based on the decisions of a single firm. Indeed, the amount supplied in a market on any particular day will be the amount that results when the outputs of all the firms in the market are added together. The point here is simply to establish the idea that there is a marginal cost of producing and offering for sale one additional unit of a good. Since the production of an additonal unit of a good requires allocating more of a society's resources for its production, under certain circumstances we can think of its marginal cost as the cost *to society* of producing an additional unit of it.

The marginal cost of producing a product differs from the *average cost* of its production, which is defined as *the total cost of producing a certain number of units of a good or service divided by the total number of units produced.* While the marginal cost is the addition to the total cost of producing just *one* additional unit at a particular level of output, average cost is a measure of the cost of producing *all* of the units being produced at any given time.

> **Average cost** is the total cost of producing a certain number of units of a good or service divided by the total number of units produced.

In many cases both the marginal and the average costs of producing a commodity *rise* as more units of it are produced (beyond some minimum level of output). This is because as production expands more of at least one input becomes more costly to obtain, and this makes it more expensive to produce additional units of the commodity. For this reason the marginal cost curves for many commodities will be upward sloping to the right on a graph that has the number of units of output (per time period) on the horizontal axis and the corresponding marginal cost for each level of output on the vertical axis.

The reason average and marginal costs may rise with increasing output is easiest to see in agriculture or other natural resource–based industries. In these cases there is a limited amount of good land (or easily exploited natural resources), so production will be more costly at higher levels of output. This is because poorer quality land—or deeper mines or more remote forests—must be used.

However, there are situations in which average cost *falls* as more is produced. Such a fall in average cost will result whenever there are *increasing returns to scale.* Increasing returns to scale exist when an increase in inputs—an increase in the *scale* of production—brings about a more than proportional increase in output. For example, we could say that increasing returns are present when doubling all of the inputs has the effect of producing more than twice as much output. Thus, as explained in Chapter 3, increasing returns produce *decreasing average costs,* and the two terms, since they refer to the same phenomenon, can be used interchangeably. To avoid confusion, however, we will generally use the term *increasing returns to scale,* and hereafter we will often just refer to *increasing returns.*

> **Increasing returns to scale** exist when an increase in inputs—an increase in the *scale* of production—brings about a more than proportional increase in output.

Situations in which firms experience falling average costs (because of increasing returns) are common and important throughout the economy. Surprisingly, however, this familiar situation—in which average costs are falling—is rarely discussed in conventional economics textbooks, and we will see why in the next chapter. But first we must carry the discussion of supply and demand a bit further.

SUPPLY AND DEMAND INTERACTING

We can now join the two strands of the story by explaining how supply and demand together will determine both the price of a good and the amount of it that will be traded (the *price* and the *quantity*). It will be useful here to refer once again to the hypothetical beer market in Iowa City and in the process to consult Figure 8.1 one more time (since it combines in a single graph both the supply curve and the demand curve for beer in Iowa City).

Of course, neither the buyers nor the sellers see the supply and demand curves. These are just analytical tools that *we* use to understand what *they* do. In most markets the sellers, not the buyers, set prices. Each seller sets a price assuming that a higher price will mean more profits per unit sold and that a lower price will mean more units sold. Depending on the demand curve, a move in either direction (that is, moving the price higher or lower) might increase the total amount of profits.

The price that maximizes profits will depend on what the other sellers are doing and how strong the demand is, two pieces of information about which a seller can guess but cannot know in advance. The one thing that sellers do know and may act on is what happened in the recent past. If goods have piled up on their shelves, they may wish that they had been charging a lower price, and they will most likely consider reducing the price. Conversely, if they sold out before the end of the day (or the month), or if they have accumulated a waiting list of eager customers, they probably will be considering a price hike. Other sellers are engaged in the same trial and error method to get the price right.

> **Excess supply** exists when at a particular price more of some good or service is supplied than is demanded.

To see how this process works, suppose, looking at Figure 8.1 again, that the average price of beer in Iowa City is $2 per bottle. What will happen? As pointed out earlier, suppliers will wish to sell 2,960 bottles at this price, but demanders will want to buy only 1,040 bottles. The difference of 1,920 bottles is referred to as *excess supply.* Those suppliers who can find buyers at $2 a bottle will be happy, while those who cannot find

buyers at this price will be dissatisfied, and the second group will then offer slightly lower prices, say, $1.75 or $1.50 per bottle, in an effort to attract customers.

As long as excess supply persists, some suppliers will cut their prices in order to try to gain customers, and this will exert a downward pressure on the market price. When the prevailing price has fallen to $1 per bottle, the quantity that suppliers wish to sell (2,000 bottles) will be just equal to the quantity demanders wish to buy (also 2,000 bottles), hence there will be no more excess supply. Price cutting by suppliers will therefore stop at this price.

> **Excess demand** exists when at a particular price more of some good or service is demanded than is supplied.

Consider the opposite situation. If the initial price in the market had been $0.50 per bottle, there would have been *excess demand.* As noted earlier, the demand for beer at this price would be 3,760 bottles, but the supply would be only 240 bottles, leaving an unmet (or "excess") demand of 3,520 bottles. The supply of beer would fall short of the demand by this amount, and those unable to buy would tend to bid up the price. Excess demand would be eliminated only when the market price reached $1, which, of course, is the same figure we arrived at in the analysis of excess supply.

> **The market clearing price** is the price at which buyers want to purchase exactly the quantity that sellers want to sell.

Summing up, we can say that competition in the market for beer pushes the whole market toward a *market-clearing price*—the price at which sellers want to sell exactly the quantity that demanders want to buy. At such a price neither excess supply nor excess demand will exist, and the market is said to "clear." As can be seen in Figure 8.1, the market-clearing price in the Iowa City beer market is $1, for at this price the quantity of beer supplied is precisely equal to the quantity demanded (2,000 bottles).

Figure 8.1 also shows that the market-clearing price and quantity are located at the *intersection* of the supply and demand curves. For this reason—and, as we shall see, only in markets similar to the one for beer in this example—it can be said that supply and demand *determine* the price and the quantity, meaning, more precisely, that the particular positions of the supply and demand curves (and, of course, the factors that themselves determine the positions of these curves) determine the market-clearing price and quantity.

> **Equilibrium** refers to a situation—a price and quantity exchanged—in which there are no forces internal to the situation pushing it to change.

In a market such as the beer market, market clearing is often described as an *equilibrium* situation, and the concept of equilibrium is important in economic reasoning. It is used to describe *a situation in which there are no forces internal to the situation pushing it to change.* This concept is borrowed from physics, and it can be illustrated with a physical example: if one drops a marble into a bowl that is sitting on a table, the marble will roll around for a while, eventually stopping at the bottom of the bowl. The result is an equilibrium, for nothing internal to the situation (the location of the marble in the bowl and the shape of the bowl) will cause it to change. If one were to tilt the bowl or push on the marble, of course the marble would move, but these would be forces external to the situation.

Economists reason the same way. In the beer market the price and the quantity sold will remain at the market-clearing price and quantity until something from the outside changes them. A change coming from the outside would be something like a change in people's taste—or, as economists like to say, their "preference"—for beer relative to other things, and such a change would bring about a movement of the demand curve DD. Similarly, the adoption of a new technology for producing beer would change the position of the supply curve SS. The resulting movements in these curves would change the market-clearing price

and the amount of beer sold. But as long as the demand and supply curves remain in their present positions, the equilibrium price and quantity will not change.

As we will see, the equilibrium price in some very important markets is not the market-clearing price. Nevertheless, the stability of an equilibrium situation (barring the appearance of any external, or *exogenous,* source of change) means that nothing an individual buyer or seller may try to do can change the equilibrium price or quantity. Thus, none of the buyers or sellers of beer in Iowa City can benefit from any possible change in their behavior, given what all the other market participants are doing.

For example, a buyer might like to pay less than the going price for a bottle of beer. But if such a buyer tried offering a lower price, no supplier would sell her or him any beer. Similarly, a supplier might like to sell beer at a price higher than the market-clearing price. But if any company raised its price, its sales would fall drastically since similar beer would be for sale in the market at a lower price and this particular company's customers would switch to other suppliers, especially if the company persisted in charging a higher price for a long period of time. (Recall that throughout this chapter we are assuming that any bottle of beer in the Iowa City market is exactly the same as any other bottle; although this assumption is somewhat unrealistic in this case, there are other commodities, such as wheat, corn, and milk, that are more like the *homogeneous* product in our example.) Thus, the prevailing market price (since it is available to anyone in the market) limits what any individual buyer or seller can do. This is the way a competitive market works.

In reality, of course, individual sellers will try changing their prices to see if they can do better. Even when there is no excess supply or demand, therefore, not all prices of a good will be the same. This may be confirmed by pricing beers at a few local stores or by checking the price of a book both at Amazon.com and on the Barnes and Noble Web site. But it is not likely that prices of the same good will differ very much if there is a high degree of competition in the market for that particular good.

An important result of the analysis of the interaction of supply and demand in competitive markets is that *when a competitive market is in equilibrium, the price of the good will be equal to its marginal cost.* Another way of putting this is to say that in equilibrium $P = MC$ (where P stands for the price and MC represents the marginal cost). We will have more to say about this idea later, but the logic of it is as follows: If P is *not* equal to MC the amount supplied will change, so the market cannot be in equilibrium. To see this, imagine that P is greater than MC for some firm. A firm in this situation can gain by producing one additional unit, increasing its revenue by P at a cost of only MC. Similarly if P is less than MC for some firm, that firm can gain by producing one unit less (reducing its costs by MC but reducing revenues by only P). So the amount supplied by a firm will not change only if $P = MC$. Moreover, for the market as a whole to be in equilibrium (recall the definition of an equilibrium) P will have to be equal to MC for every firm in the market.

SHIFTS IN DEMAND OR SUPPLY

So far we have been considering how price and quantity are determined when the supply and demand curves are in a particular position. We have looked at each curve and asked how the quantity demanded or the quantity supplied would change in response to a certain
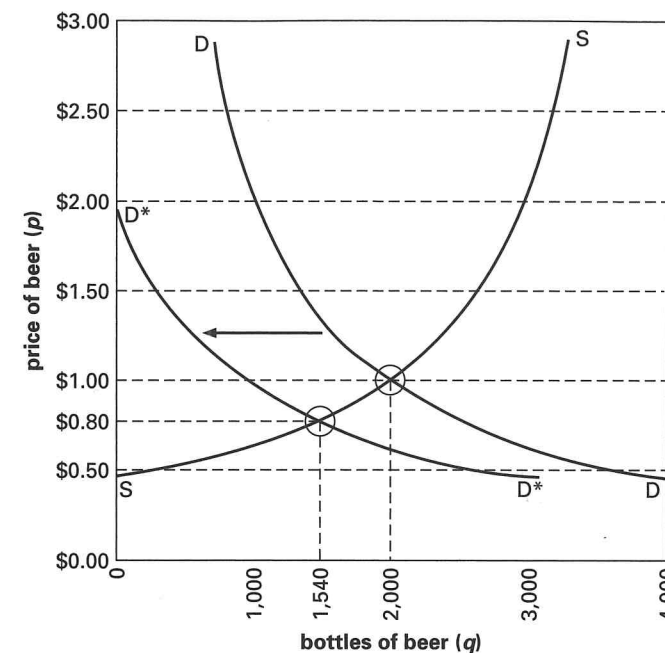
FIGURE 8.2   A shift of the demand curve.   When something other than the price changes, there will be a *shift* of the demand curve, either to the right or to the left. A shift of the demand curve is different from a *movement along* the curve, which is what happens when only the price changes. In this figure, D*D* shows the position of the demand curve after it has shifted to the left, indicating that there is now less demand for beer at every price. It is also possible that, with a different change in one of the influences on the demand curve, it would shift to the right.

change in the price. We will now consider what happens if something else besides the price changes.

For example, suppose that the demand curve DD in Figure 8.2 represents the demand for beer in the middle of a certain semester at the University of Iowa. As the semester comes to an end and students leave campus for home, the situation will change, and smaller quantities of beer will be demanded in Iowa City at every possible price. This change is represented by a leftward *shift* of the whole demand curve from DD to D*D* (see Figure 8.2). On the other hand, an effective advertising campaign during the semester would have the opposite effect: it would shift the demand curve rightward by bringing about an increase in the consumers' preferences for beer.

The difference between the effect of a change in price and the effect of something that changes the position of a demand curve can be understood as follows: a change in the price alone produces a *movement along* the demand curve, whereas a change in one or more of the conditions underlying the demand for the product produces a *shift* of the demand curve. As can be seen in Figure 8.2, the shift of the demand curve from DD to D*D* changes the market-clearing price from $1 to $0.80 per bottle and the quantity sold from 2,000 to 1,540.
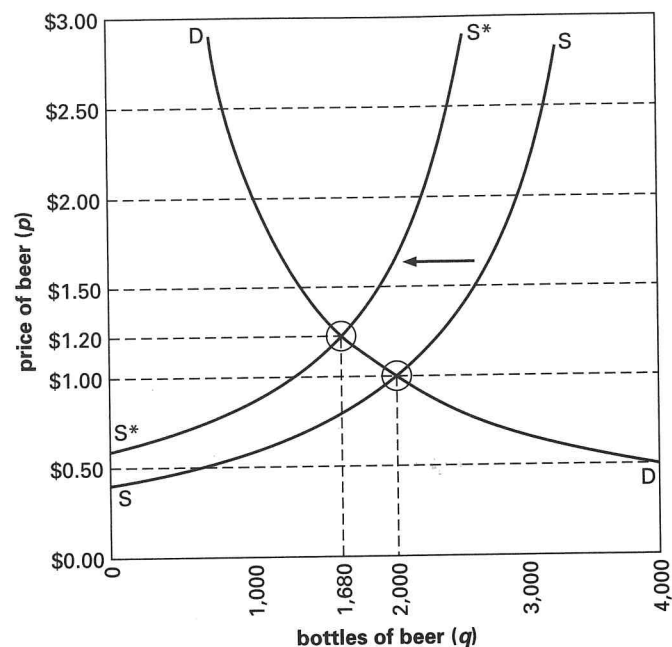
FIGURE 8.3   A shift of the supply curve.   When something other than the price changes, there will be a *shift* of the supply curve, either to the right or to the left. A shift of the supply curve is different from a *movement along* the curve, which is what happens when only the price changes. In this figure, S*S* shows the position of the supply curve after it has shifted to the left, indicating that there will now be a smaller amount of beer supplied at every price. It is also possible that, with other changes in one of the influences on the supply curve, it would shift to the right.

Similarly, the supply curve for beer will shift if there is a change in something affecting supply other than the price. Suppose, for instance, that the price of the grain used in making beer rises. What will happen? The additional cost will reduce profits in the beer industry, and this will lead some suppliers to withdraw from the beer market, perhaps to utilize their resources in a more profitable industry. The higher costs will also cause other beer producers to scale back their operations in the hope that this will restore their previous level of profits. These two developments will shift the whole supply curve from SS to S*S* (see Figure 8.3). As can be seen in Figure 8.3, this shift in the supply curve, with DD unchanged, will cause the market-clearing price to rise from $1 to $1.20 per bottle and the quantity sold to fall from 2,000 to 1,680.

As with the demand curve, a change in the price of the good creates a movement along the supply curve, whereas a change in something else affecting supply causes a shift of the entire curve. Shifts of a supply or a demand curve occur whenever one of the determinants of these curves' positions changes—whether it is, in the case of our beer market example, a

TABLE 8.1 Determinants of the Positions of Supply and Demand Curves	
The Supply Curve	**The Demand Curve**
Technology	Consumers' tastes or "preferences"
Costs of inputs, including the costs of obtaining the necessary permissions to use patented ideas, copyrighted material, etc.	Consumers' incomes and their distribution: more income at the high end will mean more demand for luxury goods; more at the low end will mean more demand for basics
Alternative opportunities for profit available to suppliers	Number of alternative products, and their prices, available to consumers
Number of potential producers	Number of potential buyers

new brewing technology, cheaper grain, a successful advertising campaign, or a change in the size of the relevant population. The determinants of the positions of supply and demand curves are summarized in Table 8.1.

CONCLUSION

Markets, then, provide a way of coordinating economies by means of buying and selling, or exchange, relationships. Competition, both among buyers and among sellers, tends to produce a market-clearing price, at which the quantity demanded is equal to the quantity supplied. The result is that market interactions determine both the prices of goods and the quantities of them that will be bought and sold.

But what does it mean to say that the equilibrium price of beer and the quantity of it that is sold are "determined by supply and demand"? This is a little like saying that a murder was committed by the gun. What actually determined the price and the quantity was whatever determined the positions of the supply and demand curves, because these determinants are what made a particular combination of price and quantity necessary to clear the market.

Supply and demand curves themselves do not do anything: they are even less involved than the gun. Rather, they are ingenious devices that help us understand and explain the many and complex influences on prices and quantities. The invention of supply and demand curves did not change the way markets work. However, it did lead to a better understanding of how markets function.

To summarize: Price and quantity are determined by all the factors that determine the positions of both the supply curve and the demand curve. The process of competition, in turn, works to make actual prices and quantities move toward equilibrium prices and quantities.

As we will see in later chapters, competition does not always cause markets to clear. In the labor market, for example, excess supply (unemployment) can persist for long

periods of time or even indefinitely. In Chapter 12 we explain why wage reductions do not have the effect of clearing the labor market—as price reductions in a beer market do—when there is excess supply. A similar situation exists in many credit markets: at the going interest rate for a given type of loan, some prospective borrowers are refused a loan, indicating that although the credit market may be in equilibrium, there is excess demand for such loans.

When markets do not clear, supply and demand will influence the price and the quantity but can hardly be said to determine them, for there is obviously some other important influence at work if the market is not clearing. Our next chapter extends the analysis of supply and demand, explaining how markets may, under certain curcumstances, coordinate the economy in beneficial ways, while, under other circumstances, they may fail to do so.

SUGGESTED READING

Alfred Marshall, *Principles of Economics,* 8th ed. (London: Macmillan, 1920).

CHAPTER 9

Competition and Coordination: The Invisible Hand

In the late 1980s and early 1990s the governments of Poland, the Soviet Union, Czechoslovakia, and other once Communist countries took a dramatic turn. Not only did they begin the process of granting voting rights and democratic freedoms to their citizenry, they also began to place more emphasis on markets rather than centralized economic planning as a way of coordinating their economies.

For more than 70 years in the Soviet Union and for 40 years in the other countries, the government made most important economic decisions. A system of economic planning collected information and then implemented decisions on such issues as how much steel and other investment goods to produce, what technologies to use, how many schools to construct, how many and what type of consumer goods to produce, and how much workers, managers, doctors, and others should be paid.

These centrally planned economies initially achieved major improvements in education, health, and other aspects of living standards, particularly for the less well-off members of their societies. But during the 1980s their governments began to implement major economic and political reforms in response to popular dissatisfaction not only with the lack of democratic rights but also with the slowing down or even reversal of economic growth. The most important economic reform was to allow private companies and individuals to make more economic decisions, and this meant relying on markets to coordinate millions of decisions. Certainly Adam Smith would have been pleased with these reforms. His idea of the invisible hand (coordination by markets) was spreading to populations whose rulers had seemed to be influenced by the single line in all of Karl Marx's writings in which he asserted that the purpose of communism was to do away with markets.

It will be decades before we can tell whether the economic reforms of the past two decades in these countries have been successful. Even then, the debate will probably continue, for the meaning of *success* will not be easy to agree upon. But the changes in economic approach in eastern Europe and the former Soviet Union dramatize the main idea of this chapter: *markets provide a way for individuals and firms to organize some aspects of their interdependence; as they do this, markets coordinate the many and complex activities that make up the economy, with no one in particular directing the process.* This idea is expressed in two main points:

---

1. By rewarding success and punishing failure, competitive markets provide a decentralized system of *motivation;* through market prices they also transmit *information* about the relative scarcity of various goods and services.

2. Under certain circumstances markets *coordinate* the economy in ways that are generally beneficial, but when the right circumstances are not present, markets fail to perform this function well.

COORDINATION

The "production-reproduction cloverleaf" presented in Figure 5.1 is like an aerial photograph of the economy. It represents the horizontal dimension of the economy as a complex circulation of goods and services—and people. Things produced in one location end up being used in another. People born and raised in the home leave and take up employment in factories, offices, or their own businesses; sometimes they work in their own or other people's homes. The circulation of goods, services, and people is called the horizontal dimension because from this perspective things and people do not move "up" or "down" in the economy. Rather, they move "across," that is, from place to place.

Understanding the horizontal dimension of the economy means explaining the movements of things and people from one place to another, answering questions such as: Why do some people work at certain jobs and some at others? What determines who and how many will raise children, pour concrete, or make shoes? Why do the resulting products move from one labor process to another? How do the shoes get from the shoe worker to the shoe wearer? Why and where do the children go to work when they grow up?

In various societies over the course of human history and around the world today, these questions have been answered in very different ways. Consider, for example, a self-sufficient family farm on the American frontier in the early 19th century. Here most of the labor processes and movements of products shown in Figure 5.1 took place *within* a single household. The production of necessary inputs as well as the reproduction of people all took place (more or less) under one roof. Tools were made and repaired, draft animals were tended, a new fence was put up, food was prepared, firewood was collected, children were born and raised, and clothing was made—all of which meant that the farm was largely *self-sufficient.* What products were produced and how they were used was coordinated by a combination of custom, necessity, and patriarchal authority. Tasks were assigned according to age and sex. Though not common today, this is one distinct way of determining who will do what labor and how the resulting output will be used.

If each family does not produce everything its members use (if the family unit is not self-sufficient)—as is the case in almost all of the modern world—the situation is much more complicated. *Specialization* will exist among, as well as within, family units, and families will have to engage in some form of *exchange* with other families. Various social arrangements will determine how labor will be specialized and how outputs will be distributed.

In some parts of India, for example, people are born into occupational groups (castes), so what one does is determined by birth. This is not as exotic as it sounds: in many parts of the world the job one does is determined by one's sex. And, as noted on the first page of this chapter, the distribution of output as well as the allocation of jobs in countries such as the Soviet Union were determined by central planners until the collapse of the Soviet-type system. Historically, there have also been some societies in which theft or tribute has played a significant role. In still other societies production and distribution have been organized almost entirely by an elaborate process of gift giving. Caste, custom, plan, gift, theft, and tribute are all ways of determining what will be produced, for whom, and by whom. They are all different methods of *economic coordination.* While many of them play a part in most economies that we know of, by far the most important methods of economic coordination in the modern world are markets and planning. We refer to these as "coordination by rules" and "coordination by command."

COORDINATION BY RULES AND BY COMMAND

Seldom in history has anyone proposed so startling a notion as Adam Smith's concept of the invisible hand. As we saw in Chapter 4, Smith held that the economy, even if it is quite complex, does not need to be run by anyone at all. Another way of putting this is to say that the economy will run itself—if it is simply subjected to the right rules.

Imagine someone telling you that an invisible hand will coordinate the landings and takeoffs of planes at O'Hare Airport in Chicago (where there are, on an average day, more than 100 landings and takeoffs per hour). The advocate of the invisible air traffic controller might say, "We don't need the control tower. Pilots should just keep right and yield to traffic on their left or below them." Given this method for coordinating air traffic at O'Hare, we suspect that most people would decide to avoid air travel to Chicago.

Even though the amount of activity at O'Hare is awe-inspiring, the interactions that occur in a whole economy, such as that of the U.S., are infinitely more complex. In the U.S., for example, the economy involves the interactions of approximately 25 million businesses, 100 million households, and 200 million adults, each of which are making decisions every day that affect at least some of the others. When the additional complexity of global economic interdependence is brought into the picture, one can easily see how daunting the task of coordination is—and how radical it was of Adam Smith to suggest that the economy need not actually be coordinated by anyone at all.

But Smith's notion of the invisible hand is not as preposterous as it sounds. Many of our interactions are, in fact, coordinated without a coordinator. Consider another traffic problem, this time for automobiles. In the United States we follow a simple rule—drive on

the right—and it does a pretty good job of coordinating the interactions of millions of drivers passing each other every day.

The point is that coordination can be achieved by either of two means: (1) with no one dictating anyone else's precise behavior, but everyone observing a set of rules, or (2) with someone (or perhaps more than one) directing the behavior of others. We refer to the first of the two means as *coordination by rules* and the second as *coordination by command*. The basic difference comes down to obeying rules versus obeying orders. A rule specifies a range of behaviors appropriate in a given situation (drive on the right) without specifying particular behaviors (where and when to drive). An order specifies a particular behavior (United flight 407 is to make a 90-degree left turn and land on runway 14A).

> **Coordination by rules** takes place when interactions are governed by general principles of behavior.

> **Coordination by command** takes place when interactions are governed by orders specifying precise behavior.

Which works better? Obviously, we need both types of coordination in our economy, by rules and by command, and the most appropriate method will vary with the situation. Adam Smith advocated the invisible hand because of his dissatisfaction with the particular kind of coordination by command that prevailed in 18th-century Britain. This involved, among other things, wage and price setting by the government and governmental creation of large monopolies such as the famous British East India Company. But even though he was generally in favor of less state intervention in the economy, Smith did specify several functions that he thought needed to be performed by the government (see Chapter 4).

We can understand how markets work if we first review two problems with coordination by command (planning), one having to do with *information* and the other with *motivation*. The individuals giving the commands (the planners) may not have enough information to do the job well, and those who are supposed to carry out their commands may have little motivation to do so. Moreover, the planners themselves may have little incentive to do the job well.

Air traffic coordination by command works well because all relevant information is available to the controllers (visually, on radar screens and on computer monitors), and the pilots have a powerful motivation to obey the controllers' orders: both their own and their passengers' lives depend on the accuracy with which they follow the commands. And, certainly, the controller has every reason to want to do his or her job well: a mistake might result in a loss of many lives as well as the loss of his job.

But in other situations neither the motivation nor the information is adequate. The central planners in the Soviet Union could not possibly have known the consumer tastes of the 300 million citizens of the country, so they could not make adequate decisions about what consumer goods to produce. Nor could they accurately determine the output capacity of each factory, mine, and office, so they could not assign production targets efficiently.

The problem in a large centrally planned economy is not exactly a *lack* of information: the consumers know more or less what they want, and the plant managers know more or less how much they can produce. So the information exists. The problem is that it is not in the right place: the relevant information is not readily available to the planners (the decision makers). Those who have the information may have an interest in keeping it from the

HUNGER VS. COMPULSION: COORDINATION BY RULES AND COORDINATION BY COMMAND

In England in Adam Smith's time, local governments were responsible for feeding the poor. In return for their food, poor people were required to work at particular jobs. This system, regulated by what were known as the Poor Laws, is an example of coordination by command.

Joseph Townsend (1739–1816), a geologist, collector of fossils, sometime physician, longtime Anglican parish priest, and prolific 18th-century writer, was one of the most severe critics of the Poor Laws. He thought it would be better to let hunger do the job of getting people to work. In 1786, a decade after the publication of Smith's *The Wealth of Nations,* Townsend wrote the following words in his essay *A Dissertation on the Poor Laws:*

Hope and fear are the springs of industry. . . . [But] *what encouragement have the poor to be industrious and frugal . . . when they are assured that if, by their indolence and extravagance* [and] *by their drunkenness and vices, they should be reduced to want, they shall* [then] *be abundantly supplied not only with food and raiment but with their accustomed luxuries. . . at the expense of others. . . In general it is only hunger which can spur and goad them on to labour. Yet our laws have said* [that] *they shall never hunger. The laws . . . have likewise said* [that] *they shall be compelled to work. But then legal constraint is attended with too much trouble, violence, and noise* [and it also] *creates ill will and never can be productive of good and acceptable service. Whereas hunger is not only a peaceable, silent* [and] *unremitt*[ing] *pressure but* [is also] *the most natural motive to industry and labor, it* [therefore] *calls forth the most powerful exertions. . . The slave must be compelled to work, but the freeman should be left to his own judgment and discretion, should be protected in the full enjoyment of his own* [wealth], *be it much or little, and* [should be] *punished when he invades his neighbor's property. By recurring to those base motives which influence the slave, and trusting only to compulsion, all the benefits of free service, both to the servant and to the master, must be lost.*

Anticipating by three centuries the arguments of some of today's advocates of "welfare reform," what Townsend was actually proposing in his *Dissertation* was a system of coordination by rules. The rules he was putting forward for consideration were: (1) you can eat only what you grow or buy and (2) you have no right to take anyone else's property, no matter how little you have or how hungry you are. The subtitle of Townsend's book was *By a Well-Wisher to Mankind.*

Source: Joseph Townsend, *A Dissertation on the Poor Laws* (Berkeley: University of California Press, 1971 [1786]), pp. 23–24. Some punctuation in the quoted passage has been slightly altered—and the words in brackets have been added—to assist the reader in following Townsend's train of thought.

THE KEY ROLE OF INFORMATION

The late philosopher-economist Friedrich A. Hayek posed the age-old challenge of how best to organize an economy—dating to before Adam Smith—as a problem of how to make the best use of information:

Which of these systems [central planning or competition] *is likely to be more efficient depends on the question under which of them can we expect that fuller use will be made of the existing knowledge. And this, in turn, depends on whether we are more likely to succeed in putting at the disposal of a single central authority all the knowledge which ought to be used but which is initially dispersed among many different individuals, or in conveying to the individuals such additional information as they need in order to enable them to fit their plans in with those of others.*

—F. A. Hayek, "The Use of Knowledge in Society," *American Economic Review*, September, 1945.

planners or in lying to them. Consumers may wish to exaggerate their needs in hopes of getting more, and plant managers may want to understate their production capacities so that they will not have to produce so much.

The difficulties with coordination by command do not end with problems of information. The *motivations* of the planners and the other economic actors may also be a problem. Unless the planners happen to be saints—and especially if they are not working within an effective form of democratic control—they may have little incentive to make decisions that benefit most of the people most of the time. Even if the planners had both the desire and the information to come up with a perfect plan, it would not be implemented unless both the plant managers and the workers had sufficient incentives to carry out the planners' orders.

Given the problems associated with coordination by command, we need to examine the other main method of coordination in the modern world, coordination by rules. Adam Smith's idea was that markets can take the place of planning as long as two rules govern the economy: competition and private property.

The way in which competition works to establish market prices (and quantities) was explained in the previous chapter. And in Chapter 6 we explained that the rule of private property means that the only ways by which one can rightfully acquire something are by labor, by purchase, or by gift (and the purchase or the gift must be from someone who rightfully acquired the object in the first place). Hence, if you want something and nobody will give it to you, your only options are to make it yourself or to buy it. You cannot simply take it. Thus, if private property prevails, if gifts meet few of our needs, and if most people are not self-sufficient, then market activities—buying and selling—will have to play a major role in the economy.

THE INVISIBLE HAND

Adam Smith was not interested in the price of beer in Iowa City. He was interested in how the British economy should be organized: should it be run by royal decree or should most economic outcomes be determined by the interactions of millions of buyers and sellers in competitive markets, with nobody in particular making the key decisions? He advocated the latter, a system of coordination by the rules of the competitive market. Since Smith's time his argument has been considerably refined, and some of its shortcomings have been clarified. The gist of it is quite simple, but to understand it at a deeper level we have to see what markets really do.

We are, however, not interested in what *particular* markets do: fish markets make fish available to consumers; housing markets make apartments and houses available. Rather, we want to know what markets do in general, especially when they are part of a system in which not only fish and apartments are marketed, but also those things that went into catching the fish and building the apartments—labor, materials, and capital goods.

As noted earlier, markets perform two important functions: they transmit economically important information and they provide the motivation to act on the information. Under ideal circumstances, then, markets address the two main shortcomings of coordination by command: they overcome the difficulties involving information and motivation.

> **Scarcity** is a relationship between a desire for something and how difficult it is to obtain. A highly desirable good that is difficult to obtain is said to be scarce.

The information markets provide is about the degree of *scarcity* of each good or service. In a competitive market the price of a good is a measure of its scarcity. If the price of a good rises (relative to the prices of other goods), we conclude that it has become more scarce; if it falls it has become less scarce. By *scarcity* we mean *both* how desirable the good or service is *and* how difficult it is to acquire. Something very desirable and even necessary, such as air, may not be scarce if it is in abundant supply. And something both costly and rare—say, the ballet *Swan Lake*

THE INVISIBLE HAND

[E]*very individual . . . employs his capital* [and] *endeavours . . . to direct* [it in such a way] *that its produce may be of the greatest possible value. . . .* [Thus] *every individual necessarily labours to render the annual revenue of the society as great as he can. He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it . . .* [H]*e intends only his own gain, and he is in this . . . led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it.*

—Adam Smith, *The Wealth of Nations* (New York: Random House, 1937), Book IV, Ch. II, p. 421.

performed by elephants—will not be scarce unless it is in great demand. Under some conditions (as we will see shortly), market-clearing prices provide a measure of both the desirability of a good to consumers and the difficulty of acquiring the good.

The motivation the market provides is of two kinds. First, markets encourage consumers to try to meet their needs with goods that are less scarce than other goods. This happens, perhaps without their knowing it, when people shop around for the best buy, seeking the good that will satisfy a particular need at the lowest available price. If, for example, certain people would be just as happy with a hamburger as with a tenderloin steak, the relative prices of the two dishes will induce them to satisfy their hunger in the way that takes the smaller toll on society's resources.

Second, the market encourages producers, either companies or individuals, to produce things that are scarce using inputs that are not so scarce. This happens because things that are scarce tend to fetch a high price, and profit-seeking firms will try to produce them with the least costly (least scarce) inputs they can find.

Thus, both consumers and producers will seek to do something very sensible, namely, to conserve what is scarce and use what is abundant. Nobody will require them to do this. They will do it simply because it is in their personal interest to do it. As Adam Smith implied, the market achieves its results behind the backs of the participants. This is the basic argument that supports his notion of the "invisible hand."

The key idea here is that the *price* of a good measures its scarcity. According to a widely held notion of scarcity, fish are scarce when there are none or only a few to be bought; apartments are scarce when the vacancy rate is low. Thus, most people think of scarcity as an issue of quantity. Nevertheless, economists insist that scarcity is best measured by price. Why is this the case?

The quantity notion of scarcity is essentially misleading. Consider the example of gasoline. When the Organization of Petroleum Exporting Countries (OPEC) decided in 1973 to limit production of oil in order to drive up prices, what happened? At first there were long lines of cars at gas pumps. There was a shortage of gas. Everyone knew that gas was scarce, and the price rose (from $0.39 a gallon for unleaded gasoline in 1973 to $1.31 eight years later). But over time the lines at the gas pumps disappeared. Though gas was still scarce in 1981, there was enough to go around because by then the higher prices had changed people's driving habits and tastes in cars, so people were buying less gasoline than before.

The quantity notion of scarcity says that something is scarce if there is excess demand for it. But what happens when excess demand exists? The price rises, and the excess demand tends to disappear. Thus, rather than focusing on excess demand, the economist takes a higher equilibrium price as a better measure of increased scarcity.

THE INVISIBLE HAND IN ACTION

The case for competitive markets—the invisible hand—rests ultimately on the claim that even if all of the economic actors behave only with regard to their own self-interest, markets can allocate scarce economic resources in a desirable way. Advocates of laissez-faire (limited government) argue that competitive markets not only address the problems of information and motivation but in so doing offer a method of coordination that is superior to central planning (coordination by command).

The beer market in Iowa City offers an example of how a competitive market can induce producers and consumers both to respond to a change in tastes and to economize on society's scarce resources. First, imagine that the beer market is in equilibrium and that the market-clearing price is $1 a bottle. Then, suddenly, the U.S. Surgeon General comes out with a report saying that drinking wine causes baldness. What will happen in the beer market?

As people switch from wine to beer, the demand curve for beer will shift to the right, indicating that there is now a greater demand for beer at each price. The result, shown in Figure 9.1, is an excess demand for beer. At the price of $1, producers are willing to supply 2,000 bottles. This means that, as before, they can make as much profit producing and
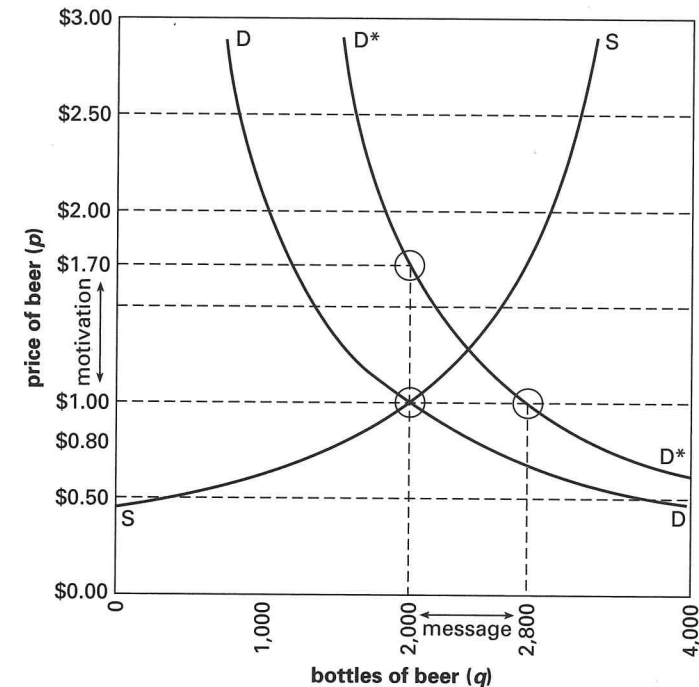

FIGURE 9.1   Market messages and market motivation.   The shift of the demand curve to the right (from DD to D*D*) creates excess demand in the hypothetical Iowa City beer market. An excess demand of 800 bottles will exist if the price per bottle remains at $1.00 because at that price the supply is only 2,000 bottles while the demand is for 2,800. Although the first 2,000 buyers will get their beer at the going price of $1.00 per bottle, there will be a line of 800 disappointed customers, none of whom will get any beer—even though some would have paid up to $1.70 for a bottle—because the supply will have run out. Thus the market is sending the *message:* "More beer!" Since consumers are willing to pay more than $1.00 per bottle, suppliers will find it profitable to provide an additional bottle of beer at any price above $1.00. Thus, at any price between $1.00 and $1.70, additional supplies (and sales), up to 2,800 bottles, will be pleasing both to suppliers and to consumers. Because suppliers can increase their profits by selling more while also raising the price, they will be motivated to increase the supply of beer. Thus, the price differential of $0.70 is the *motivation* provided by the market. It says: "Producers who pay attention to the market message will make more money."

selling 2,000 bottles of beer at that price as they could doing anything else with the resources available to them. But if, at that price, they were to produce only that quantity, there would be an excess demand of 800 bottles (since the demand for beer at a price of $1 is now 2,800 bottles). $1 is obviously no longer the market-clearing price, and the market (in the form of excess demand) is sending the beer producers the *message* that more beer could be sold without lowering the price.

Now look at the excess demand another way, not as a quantity, but as a price. Notice that in the situation in which only 2,000 bottles are available (that is all the producers are willing to supply at a price of $1), some consumers are willing to pay $1.70 a bottle (this is the price the new demand curve shows that consumers would pay for that quantity). The difference of $0.70 between the two prices provides crucial information: the price consumers are willing to pay for another bottle of beer is greater than the cost producers would have to pay to supply another bottle. It is obvious that there is some price between $1.00 and $1.70 at which an increase in the amount of beer supplied would please both consumers and producers. From this economists would conclude that more beer in the market would benefit all concerned.

How, then, does the market provide the *motivation* to produce more beer? If the producers notice the presence of excess demand—how could they miss it?—they will also realize that they can raise their price and still sell at least the 2,000 bottles they are now producing. But once they have raised their price, they will be making more profit brewing beer than they could make in any other activity. (Among the assumptions about competitive markets made in this chapter is the notion that in equilibrium the level of profit in each industry will be the same as the level in every other industry, otherwise the economy would not be in equilibrium since there would be a tendency for some entrepreneurs to shift their resources from a low-profit to a higher-profit industry. Hence, when the beer industry was in equilibrium, the market-clearing price of $1 per bottle would allow beer producers to make just the level of profit that was being made in every other industry.)

Accordingly, the motivation to expand the production of beer comes from the opportunity now available to beer producers to make more profit by selling beer at a higher price. The higher rate of profit per bottle of beer will induce the beer suppliers to produce more beer. This is, of course, just another way of describing the process through which the market-clearing price and quantity are arrived at. However, this analysis leads to an important conclusion: the profit seeking of the beer suppliers will not only lead them to do what is in their own interest, it will at the same time lead them to do what is in the interest of consumers. Moreover, insofar as this process leads to the best possible use of the economy's resources, it may also be said to promote the interests of society as a whole.

To summarize: If excess supply exists, the market provides producers with motivation in the form of a *stick:* firms must either adapt to the information the market is providing—by producing less—or go out of business. When excess demand exists, the market offers producers a *carrot:* higher profits await those who grasp the meaning of the market's information and expand production. In this way the market directs self-interested producers and consumers to do what is in the interests of both, even if neither cares about the well-being of the other. This is the invisible hand in action.

How did the invisible hand overcome the two main shortcomings of coordination by command, namely, the planners' lack of relevant information and the absence of incentives either to plan well or to carry out good plans should they happen to occur? The answer is

TABLE 9.1	Prices Are a Decentralized System of Information and Motivation

Prices as *information*

- to *consumers:* the price measures how much it costs to produce an additional unit of a commodity.
- to *producers:* prices measure how much demand there is for an additional unit of the commodity and how much it costs to acquire the necessary inputs.

Prices as *motivation*

- to *consumers:* prices, in conjunction with the need to stay within one's budget, motivate consumers to satisfy their wants as cheaply as possible.
- to *producers:* prices, in conjunction with the need to make money in order to stay in business, motivate the lowest-cost production of goods and services that consumers want.

TABLE 9.2	The Invisible Hand: Assumptions and Conclusions

If the prices of goods, as they are sold to consumers, measure the ability of the goods to satisfy human needs, and

If the costs of producing goods, as measured by firms, take into account the social costs of acquiring and using the goods,

Then the profit made on each unit of a good (the price minus the cost per unit) will measure the social contribution made by producing each good, and

Hence the pursuit of self-interest (firms seeking greater profits and consumers trying to maximize their satisfactions) will result in a socially desirable allocation of our human and natural resources.

decentralization. Prices provide information about consumers' wants and producers' capacities and costs, and this information is communicated directly among all the market participants without first having to go through a central planning office. Also, the incentives to "do the right thing"—with regard to the efficient allocation of resources—are based simply on the self-interest of all the participants. When all the players in the market seek the best possible deal for themselves, everyone benefits.

If all this sounds a little too good to believe, it is because we have not yet asked how the invisible hand argument might work (or not work) if the ideal circumstances we have been assuming are not, in fact, present.

> **Coordination failure** occurs when markets or other types of coordination by rules fail to coordinate an economy in such a way as to produce outcomes that are desirable.

PROBLEMS WITH THE INVISIBLE HAND

There are many situations in which markets do not perform so well. When markets fail to coordinate an economy in such a way as to produce outcomes that are desirable, economists call this a *coordination failure.* Two parables can be used to introduce the concept of a coordination failure. The first is "the prisoner's dilemma," and the second, "the tragedy of the commons."

The Prisoner's Dilemma and the Benefits of Cooperation

This widely circulated story makes the following point: individuals who *cooperate* may come closer to achieving their personal objectives than do those who pursue their self-interest without regard to the well-being of others. The theory of how markets coordinate an economy—conveyed in Adam Smith's metaphor of "the invisible hand"—shows that *under certain conditions* competition based on self-interest but coordinated by markets will bring about a desirable allocation of economic resources. The story of the prisoner's dilemma, however, dramatizes the fact that *under other conditions* lack of cooperation results in a coordination failure.

The story goes as follows. Two individuals suspected of committing a crime are picked up by the police and placed in separate cells. The police have not been able to gather sufficient evidence to convict either of the prisoners (who we will refer to as prisoners "A" and "B"), so the local prosecutor must attempt to get at least one of the prisoners to confess to the crime and implicate the other one. The prosecutor offers each of the prisoners the choice of confessing to or denying involvement in the crime, and the consequences of all the possible choices are explained. Each prisoner, in turn, must decide independently on the best strategy to pursue.

The possible outcomes are as follows:

1. If A denies involvement in the crime and B does the same, they will both be convicted of a lesser offense (regarding which the prosecutor has ample evidence), and they will each receive sentences of two years in jail.
2. If A denies involvement in the crime but B confesses, implicating A, A will be sentenced (on the basis of the evidence provided by B) to four years, and B will go to jail for only one year.
3. If A confesses, implicating B, but B denies involvement in the crime, B will be sentenced (on the basis of the evidence provided by A) to four years, and A will go to jail for only one year.
4. If both prisoners confess to the crime, they will each be sentenced to three-year terms.

These four options are summarized in Table 9.3. The shaded cells in it are numbered in accordance with the four options presented above.

TABLE 9.3 Prisoner's Dilemma

		Prisoner B	
		Confesses	Denies
Prisoner A	Confesses	4 — A and B both get 3 years	3 — A gets 1 year B gets 4 years
	Denies	2 — A gets 4 years B gets 1 year	1 — A and B both get 2 years

First, assume that both prisoners make their choices with regard only to their own interests. What choices will they make? It will become apparent that under this assumption each prisoner can do better—*whether the other prisoner confesses or denies*—by confessing to the crime and implicating his or her partner.

Imagine that you are prisoner A. If you deny involvement in the crime while your partner is confessing (implicating you), you will be put away for four years (shaded cell 2). If, however, you confess while your partner is confessing, you will get only three years (shaded cell 4). So in the case that B has confessed you (A) would be better off if you confessed. If, on the other hand, you deny involvement while your partner is denying, you will get two years (shaded cell 1), but you will get only one year if you confess while your partner is denying (shaded cell 3). In this case, too, you are better off confessing. It is evident that confessing to the crime will leave you better off whatever your partner does—hence self-interest will lead you to confess. And if your partner is self-interested, too, he or she will come to the same conclusion. In this situation, then, you will both confess to the crime, and you will both be put away for three years (shaded cell 4).

Now imagine (a) that both you and your partner care enough about each other to want to keep the other from being put away for four years, and (b) that you are both willing to adhere to this objective even though it might not seem to serve your own interest. These assumptions mean that *cooperative* behavior will prevail. In this case neither of you will confess, and both of you will serve two years in jail (shaded cell 1). This outcome is clearly better for each of you than the one resulting when you both act with regard only to your own interest (shaded cell 4). The preferred outcome would also occur if both you and your partner are people who keep promises (whether or not keeping them seems to be in your own interest) and if, before committing the crime, you each promise to deny involvement if arrested.

The moral of the prisoner's dilemma story is not that it is a good thing for criminals to serve as little time as possible in prison (whether or not they have actually committed a crime). The moral is that in some situations the pursuit of self-interest by all parties leads to outcomes in which none of the participants benefit. These situations are just the opposite of the situation described by the invisible hand.

The prisoner's dilemma can help us find answers to such questions as: why is it so difficult to prevent destruction of the environment or to establish minimum standards of workplace health and safety among nations? In these cases the choices are not to confess or deny, but to "adhere to environmental (or workplace) standards" or "violate the standards."

The prisoner's dilemma and global warming. There is a major obstacle to the progress of efforts to slow the pace of global warming: the government of any particular country would probably like to see all *other* nations limit *their* emissions of greenhouse gases (primarily carbon dioxide) while avoiding the politically unpopular task of limiting its own citizens' freedom to pollute. Many people in the rich countries resist any restraint on activities, such as driving a car, that pollute the environment; at the same time, many people in the poor countries of the world believe that limiting greenhouse gas emissions in their countries would hamper their attempts to industrialize and raise their living standards to those now common in rich nations.

The countries of the world thus face a coordination problem: if each nation's government acts independently and seeks to serve only the perceived interests of its own citizens

TABLE 9.4 North's Action/South's Action

		North's Action	
		Nothing	Protect
South's Action	Nothing	Third best for both	Best for S Worst for N
	Protect	Best for N Worst for S	Second best for both

(or its most powerful ones), no nation will adopt the measures necessary to avert major climatic change. To see why this is so, imagine that there are just two countries, called North and South, and just two options, called Nothing (meaning do nothing to prevent global warming) and Protect (meaning adopt protective measures to slow or halt global warming). Each country would like the other one to reduce its greenhouse gas emissions while avoiding the costs of doing so itself. Thus, for each the best alternative is to do Nothing while the other does Protect. The worst alternative for each is to Protect while the other does Nothing. The second-best alternative for both is that both Protect, and the third-best alternative for each is that both do Nothing.

The options available to the players in this simple game are represented in Table 9.4. This is a prisoner's dilemma situation, and as we have seen earlier, what is best for each is worst for all. If each country makes its decision independently and they both do what their own people see as being in their own self-interest, both countries will do Nothing. (As noted earlier, what is best for each does not depend on what the other does.) They would both be better off, however, if they both chose to Protect.

The best outcome can be achieved, however, only if the two countries agree to Protect, perhaps accomplishing such an agreement by signing an international treaty committing themselves to the implementation of protective measures. But two problems immediately arise. First, who will enforce the agreement? There is no world government or any other body that could force each nation to abide by its agreements. The second problem is that different ways of protecting the global environment result in differing costs and benefits for each country. Why, for example, would the poor countries of the world agree to pay an equal share of the costs of reducing greenhouse gas emissions, most of which are now caused by the high-income countries?

The poor countries might argue that each nation should have the right to pollute in proportion to its population. In this case any solution to the problem would require drastic reductions in carbon dioxide emissions in the rich countries (where pollution per capita is high), while leaving the poor countries relatively free to industrialize in a way that increases pollution. While this solution may seem fair to many people, it is likely to be politically unpopular in the rich countries.

Extending the lesson of the prisoner's dilemma more generally to economic systems, we can say that reliance on competitive markets may fail to coordinate an economy in a desirable way: the pursuit of self-interest by individuals may result in outcomes that are not optimal from the point of view of any member of the society. This is what is meant by the term *coordination failure*.

The Tragedy of the Commons

Another illustration of a coordination failure is referred to as "the tragedy of the commons," a parable set forth in 1968 by Garrett Hardin in a now-classic article in *Science* magazine.[1] This parable is about the possibility that environmental destruction will result from the uncoordinated pursuit of individual self-interest.

Imagine a large lake, its shores dotted by the houses of people who fish in the lake to earn their livings. No one owns the lake: it is the common property of all the members of the surrounding community. The situation described here, one that involves communal ownership of an important shared resource, is similar to arrangements that were widespread in many of the early settlements in New England (it has, of course, also existed in other parts of the world at various times in history), and it takes its name from the shared grazing land for cattle and other livestock that was referred to as the town "common."

In the parable of the lake, each person decides independently the number of hours to fish each day (or, alternatively, how many fish to catch) before heading back to shore. As self-interested people, they fish as long as the additional benefit of another hour (or another fish) is greater than the inconvenience or discomfort of the additional time and effort spent fishing.

However, as in the case of the prisoner's dilemma, what is rational for one is not beneficial for all. The more each person fishes, the more difficult it is for the others to catch fish. The reason for this is simple: there are a limited number of fish to be caught, and as each person catches more of them, fewer are left to be caught by the other lakeside residents. Each person would like to see limits put on the fishing of the others while remaining free, herself, to fish without limit. As long as there are no limits placed on anyone's fishing, there will be overfishing. The end result is predictable: there will soon be hardly any fish left in the lake, and, eventually, the lake will be entirely "fished out." That is the tragedy.

This story is about fishing in a lake, but it might just as well be about overgrazing open pastureland, dumping sewage into a river, or polluting the air. The moral of the story is that the pursuit of individual self-interest can be highly irrational because it can have very negative, possibly irreversible, consequences.

A real-life example of the tragedy of the commons is the case of what happened with a particular effort to harness geothermal energy for the generation of electricity. A potentially cheap and environmentally clean way of generating electricity involves tapping the steam that shoots in geysers out of the earth. In the years since this method of producing electric power became practical, a huge field of geysers 70 miles north of San Francisco has been regarded as a promising place to try it out. When energy prices suddenly rose in the mid-1970s, the geysers north of San Francisco became the focus of intense but largely uncoordinated energy development. However, as more and more electricity-generating plants were built there, the amount of steam available for earlier-installed plants fell. (In

---

[1] Garrett Hardin, "The Tragedy of the Commons," *Science*, no. 162, Dec. 13, 1968.

such a field of geysers there is a common but limited amount of steam deep beneath the earth's surface.) Eventually, so much steam was being tapped that none of the electricity-generating plants could operate at full capacity. By the early 1990s it became clear that because the number of plants had not been limited, the geysers were no longer an efficient source of energy.[2]

How can tragedies like this be avoided? There are many answers to this question, but all involve finding some way to take others' interests into account when making decisions. The most obvious solution to the problem would be simply to regulate access to the common resource. In the fictitious case described above, the fishing people around the lake could have met, ascertained how many fish might reasonably be taken from the lake each week, and decided collectively on a fishing limit for each member of the community that would have allowed the weekly target to be met. In fact, there are fishing villages in Japan, Turkey, and elsewhere that engage in this practice.[3]

If, on the other hand, the problem was overgrazing, those with the animals could figure out how much grazing the common land could support and then decide on the number of animals each herder would be allowed to graze. Solutions to the problem such as the ones mentioned so far all involve what is termed *social regulation of the commons.*

An alternative, which might be termed the *private property solution,* would be for the commons to be owned by a private individual or company. Such an owner would, say (in the cases mentioned above), not only employ the lake residents to fish or the animal herders to graze their livestock, but also would, in self-interest, limit the use of the common resource in order to prevent its deterioration. Obviously, the owner would not want to see the lake overfished or the pasture overgrazed. A problem with this "private property solution" to the tragedy is that the common resource is often too large to be owned by a single individual or company. Something like the tragedy of the commons is currently being played out in the ozone layer high in the atmosphere above the earth, and the world's oceans are a similarly threatened common resource. For such large commons, private ownership is obviously not possible, so the social regulation approach is all there is. However, as noted earlier (see "The Prisoner's Dilemma and Global Warming" above), social regulation is not easy to work out in practice.

Market Failure

> A **market failure** is said to take place when the spontaneous interactions of buyers and sellers on markets each pursuing their own objectives results in outcomes that are generally undesirable.

The term *coordination failure* refers to *any* situation in which the self-interested behavior of individuals results in an outcome that is less beneficial to them than one that might have been achieved by better-coordinated, or cooperative, behavior. The term *market failure,* on the other hand, refers to the specific type of coordination failure that happens because of how *markets* work. Neither the prisoner's dilemma nor the tragedy of the commons, much as they may shed light on the problems

associated with (self-interested) market behavior, are in themselves examples of market failure (since they do not involve *exchange* relationships).

Why, then, does the invisible hand of the market fail, in many situations, to coordinate economic activities in ways that produce desirable results? The general answer is that people affect one another's well-being in countless ways, and market prices often fail to take into account all the effects of our actions on others. In small groups, such as families and friendships, we typically consider the costs and benefits of our actions not only with regard to ourselves but also as they might affect the others we care about. Such consciousness, also known as altruism, is an essential part of social life, but it obviously does not prevail when we are interacting with large groups of strangers, about whom we may know little and care less.

The underlying assumption of the invisible hand is that *if prices are right* (in the sense that they measure the true scarcity of all aspects of human interactions), all the effects of people's actions on one another will be taken into account in the prices of goods and services. Thus, Adam Smith's theory (including its modern version) holds that—if we assume (as Smith did) that the existing distribution of wealth is acceptable—coordinating an economy with markets will bring about an optimal allocation of society's resources.

Proponents of the invisible hand theory offer examples such as the following to support the theory: When a certain consumption choice, say, ordering a tenderloin steak, uses up a lot of society's resources, that consumption choice will be appropriately paid for in the price of the steak. This is because the price will be equal to the marginal cost—hence the consumer will be paying exactly the amount it cost to produce that tenderloin steak. Proponents of the invisible hand theory would also argue that the marginal cost is a good measure of the amount of other goods—vegetables for vegetarians, for instance—that might have been produced with the same resources that went into producing the steak.

Another example of optimal pricing would be if a person's uniquely skillful work went into producing a product that brought great joy to others, such a product would fetch a high price (reflecting its marginal cost as well as the demand for it), and its producer would be handsomely rewarded when the product was sold.

The problem is that *prices are often not right* (in the sense just described). There are many examples. If the technology used by a particular company helps it to produce a valuable product at low cost but also gives off a high level of airborne toxins, the company may be rewarded with high profits, but it does not pay for the pollution it imposes on others. In this case the company does not pay for the clean air it uses up, it pays only for the inputs it has to buy in the market. As a result, the price of the product, although it may be equal to the marginal cost *to the company* of producing the product, does not take into account the additional costs that production of this product imposes on others.

Another example: If an individual develops a new computer software application that proves exceptionally valuable to its users but is costless to copy, the "right price" (the price that is equal to the marginal cost of producing additional copies of the application) is zero. This price will obviously not allow the developer to be appropriately rewarded, and it will certainly not provide an incentive for others to develop valuable software. One might think that patent protection would solve this problem—and indeed it might from the point of view of the individual developer—but the enforcement of this legal right (assuming it could be enforced) would not result in an optimal allocation of resources. This is because

---
[2] Richard Kerr, "Geothermal Tragedy of the Commons," *Science,* no. 253, July 12, 1991, pp. 134–135.
[3] F. Berkes, D. Feeny, B. J. McCay, and J. M. Acheson, "The Benefits of the Commons," *Nature,* no. 340, July 13, 1987, pp. 91–94.

if others are prevented from using the application without paying its developer a substantial price, the price charged would be much higher than its marginal cost of production.

A final example: If you spend many hours of your time at local school board meetings or doing volunteer work to improve the quality of life in your neighborhood, you will undoubtedly reap personal benefits from these activities. However, there will not be a market price—or payment—that compensates you for any of the benefits received by your neighbors. Moreover, if you are a renter of your house or apartment, there will be no payment rewarding you for the resulting increase in the value of your residence, since it belongs to your landlord. Indeed, your landlord might raise the rent she is charging you, arguing that the market values (and rental prices) of houses and apartments in your neighborhood have gone up as a result of the improvements in the neighborhood and its schools.

Why do markets fail in these cases? One reason is that in these cases prices do not adequately measure the scarcity of the goods in question. The product of a polluting factory is more scarce than its low price seems to indicate because the cost to society of the factory's pollution is not included in the price of the good. The computer application, once produced, is not scarce at all: it can be copied over and over again at zero marginal cost. Hence, if its developer somehow manages to sell it at a profit, its price will overstate its scarcity. But if it were to be sold it at its true scarcity price (zero), it would have to be given away, in which case there would be no incentive to produce it in the first place. Finally, the volunteer activities of citizens are, by definition, not paid for, so it is impossible to put market prices on such activities even though the supply of community-oriented labor is a scarce as well as a valuable resource from the standpoint of other citizens and society as a whole.

> **Private costs** are the costs borne by the user of a good or service (a person or a company), while the total costs borne by all members of a society are termed **social costs.** At best, prices only measure private costs.

In all these cases there is a discrepancy between (a) the costs and benefits borne or received by the decision maker and (b) the costs and benefits experienced by all the members of society (including the decision maker). The costs and benefits accruing to the decision maker are referred to as the *private costs* and benefits of the activity in question, whereas when added together all the costs and benefits experienced by everyone (again including the decision maker) are termed the *social costs* and benefits.

When private and social costs diverge, economists say (as we explained in Chapter 3) that there is an *externality* because individuals or groups "external" to a transaction experience some of its effects. Another way of putting this is to say that there are *spillover* effects if some of the effects of a transaction "spill over" and confer benefits or impose costs on individuals or groups not directly involved in a transaction. We favor the use of the term *externality* because it explicitly distinguishes people who are "internal" to—or directly involved as decision makers in—a transaction from those who are "external" to—or not direct participants in—the transaction.

When there is an externality it can be either positive or negative, positive if the external effects (or spillovers) of a transaction are beneficial to others, negative if they impose costs on others. Thus, there are positive externalities and negative externalities. In addition, however, positive externalities are sometimes referred to as *external economies,* while negative externalities are termed *external diseconomies.* How confusing! We continue to favor the use of the simpler term *externality*—as in positive externality and negative externality.

An important consequence of externalities is that when they exist, private marginal costs are not equal to social marginal costs. In this case even if prices are equal to private

TABLE 9.5 Positive and Negative Externalities

Action	Externality	Price Not Equal to Social Costs/Benefits
Drinking too many beers	The costs of alcohol abuse on families, friends, and medical providers	The price of the beer does not reflect such negative externalities
Driving a car	This can contribute to traffic congestion, and it also causes environmental degradation from both the production and the use of fossil fuels	The price of the gas used does not reflect these negative externalities
Research	The benefits of any good ideas produced will most likely be enjoyed by other people or companies	Any increase in the profits of the firm that supported the research—or rise in the salary of the individual who conducted it—may not take into account all the benefits of the resulting knowledge enjoyed by others
Education	The benefits of one person's education are enjoyed by neighbors and workmates	The individual benefits (higher earnings) do not capture the positive externalities enjoyed by others
Training employees	Some employees will move to other firms, the owners of which will benefit from the training paid for by the firm that did the training	The training firm's profits do not reflect the benefits flowing to other firms with the workers who move
Wearing a $500 watch (luxury consumption)	This may have the effect of lowering the (relative) status of others, creating envy	The $500 price of the watch does not include the status and envy costs imposed on others

marginal costs, they will not be equal to social marginal costs. Hence, they will not measure the cost to society of producing an additional unit of a good.

Important sources of positive externalities are education, the production of knowledge, and the introduction of new technologies. Important examples of negative externalities are pollution, other kinds of environmental degradation, and automobile traffic congestion in metropolitan areas. Some examples of positive and negative externalities are presented in Table 9.5.

Another reason why externalities exist in competitive markets is that while some costs and benefits are covered in contracts, others are not. As noted in Chapter 3, this is the problem of *incomplete contracts.* When someone sells a commodity to another person, the buyer pays the seller the contracted price. But when there are social costs involved—such as the costs of dealing with the consequences of pollution or enforcing the provisions in a contract—the contracted price does not cover the costs of health care, cleaning up toxic wastes, or paying other expenses involuntarily imposed on the buyer, the seller, or other members of society.

The problem of incomplete contracts can be illustrated using the examples already mentioned. In the case of software, even if there is a contract saying that copies of it cannot

be made by—or distributed to—persons other than the purchaser (think of the printed seal that is affixed to the envelope enclosing a CD containing a just-purchased piece of software), such a contract is virtually (no pun intended) impossible to enforce.

If someone's volunteer efforts have the effect of improving the quality of a neighborhood or its schools, there is no contract enabling that person to collect from the local landlords or homeowners any portion of the resulting increase in property values. And neither is there any way for the volunteer to be directly compensated by his or her neighbors for the improved quality of the neighborhood or its schools.

Another example of an incomplete contract is the wage bargain negotiated between employer and employee. The employer contracts (at least implicitly) to pay the employee a certain wage rate per hour on the job, and the employee, in return for the wage, agrees to come to work for a specified number of hours per day and for a certain number of days each week. But the amount of work actually to be done by the worker is not—and cannot be—covered in the contract. Ensuring that the employee does enough work to make the enterprise profitable is the basic problem of *management* that faces every employer. How hard the employee works, the "intensity" of his or her labor, will depend to some degree on his or her "work ethic," but it can also be affected by the type of incentives offered by the employer in return for hard work (the carrot) or the degree to which the employer can coerce the worker (by threat of firing, use of tough supervisors, etc.) into working hard (the stick). The condition of the worker at the end of the day, another variable not covered in the contract, will depend on the quality of the work environment provided by the employer and the way in which the employer exercises his or her managerial powers to get the worker to work. The challenge to employers of getting workers to work hard is discussed at length in Chapters 12 and 13.

In addition to the problem of incomplete contracts, markets may also fail for the more obvious reason that there are too few competitors, or even potential competitors, in the market. The invisible hand argument assumes that markets are competitive, with many actual or potential buyers and sellers in every market. Given this assumption, no producer can charge more than the marginal cost of producing a particular commodity.

In the real world, however, many markets are not competitive. In the case of a *monopoly,* a single producer with no actual or potential competitors can raise the price buyers must pay for its product without fear that its customers will switch their patronage to another supplier. A monopoly has *market power* since it can raise the price of its product merely by producing (and selling) less of it. This will create an artificial scarcity in the market for the product and drive up its price. In this situation the price charged will be higher than the marginal cost and hence will not be consistent with an optimal allocation of society's resources.

Of course, if there were other firms that could produce the same product, they would enter the market, would compete for customers by charging lower prices, and would continue doing this until the price of the product fell to its marginal cost. It is this process that leads to the result that in a competitive market, the market-clearing price will be equal to the marginal cost ($P = MC$) in every firm. But with barriers to the entry of new firms, which is what makes a monopoly a monopoly, a monopolist does not have to fear competition from new firms. The workings of imperfectly competitive markets are discussed at length in Chapter 11.

Prices often exceed marginal cost for yet another reason, one that has nothing to do with monopoly. To see this, ask yourself what the marginal cost was of the last CD you

purchased. If the album was relatively popular and the CD you purchased was, say, the 423,589th one produced, its marginal cost was probably not much more than a dollar. This is considerably less than its average cost of production because, as explained in Chapter 8, marginal cost is just the increase in the total cost resulting from the production of one *additional* CD, whereas average cost takes into account not only the cost of producing the additional CD but also what the CD-producing company had to pay for such things as advertising, permission to use copyrighted material, rent for the use of the production facilities, interest paid to the bank for loans, and any other expenses that do not vary as a result of producing one more CD. Economists refer to such costs as *fixed costs*. They are included in the total cost but are spread out over the total number of units produced, which is one reason why average cost (total cost divided by the number of units produced) declines as more units are produced. In the music industry and other parts of the economy such as book publishing, fixed costs are called *first copy costs*.

Of course, there are also costs associated just with the production of the one CD you purchased. The company that produced it had to acquire a blank CD, burn the music onto it, purchase its jewel case, print and install the tray card, and wrap the final product in cellophane. But when the production run is large the costs of all these steps will be very low for each individual CD because things such as blank CDs and jewel cases can be purchased in large quantities for only a few cents each. And the process of burning music onto a CD, if it is repeated thousands of times, is also very cheap. Even if the costs of paying royalties to artists and distributing the CDs to retail outlets are taken into account, the marginal cost per CD will still be very low.

Since the marginal cost of your CD was probably in the neighborhood of $1 while the price you paid for it was most likely $14 or more, the CD market is obviously not working well: it is not setting a scarcity price ($P = MC$) on the CD. There are almost certainly quite a few people out there who felt they could not afford to buy the CD at $14 but who would have purchased it—and enjoyed listening to it—had the price been $1 or $2. The existence of a number of frustrated buyers who would have been willing to pay the cost of the resources required to make available an additional copy of the CD means that *the invisible hand is not working* in this case. Why is this?

CDs are manufactured and sold by hundreds of companies, so the gap between their prices and their marginal costs of production is not due to monopoly power. Rather, the gap is probably the result of a phenomenon (discussed in Chapter 3) that prevails in many parts of the economy: *increasing returns to scale*.

As noted earlier, the presence of increasing returns means that the *average cost* of producing a product will *fall* as the quantity produced rises. And if this is the case, *the marginal cost of production will be below the average cost*. We know that this is true because of the mathematics of the situation: if average cost is rising, the marginal cost must be above the average cost (in order to be pulling it up), while, for the same reason, marginal cost must lie below average cost if the latter is falling (in order to be pulling it down).

To illustrate this point with a concrete example, assume that the total cost of producing 100 CDs is $1,005, which means that the average cost of producing the 100 CDs is $10.05. (Recall that the average cost is equal to the total cost divided by the number of units produced.) If the CD-producing company now decides to produce 101 CDs and finds that the total cost of producing that many CDs is $1,010, the average cost of producing the

101 CDs will have dropped from $10.05 to $10.00. To calculate the marginal cost in this example, we compare the total cost of producing 100 CDs ($1,005) with the total cost of producing 101 CDs ($1,010), finding that the marginal cost of producing the 101st CD is $5. Since $5 is less than $10, it is clear that when average cost is falling (as it is in this case), the marginal cost will be less than the average cost.

Returning to our example of a CD that is being produced in very large quantities: if the retail price of the CD (say, $14) exceeds the marginal cost by a large amount (say, by $13 per CD), why are new firms not entering the industry, expanding the production of CDs, and driving the retail price of them down (as the invisible hand theory of competitive markets would predict)? The answer is that the market price may be above the *average cost* only by something like $3, thereby allowing the CD-producing company to make a profit of only $3 on each CD produced. If this provides the company with a rate of profit that is not out of line with the rate of profit being made in other industries—in other words, if it is roughly equal to the average rate of profit in the economy as a whole—other firms will not have an incentive, say, to stop making frisbies in order to enter the CD-producing industry. (We provide a precise definition of the rate of profit and an analysis of how it is determined in Chapter 10.)

Although there is, as we have demonstrated, a market failure in this case, it is important to note that the CD-producing company can hardly be expected to sell its CDs at their marginal cost: to do so would mean making losses rather than profits. Rather, the problem arises because of the *system* of profit making and pricing, not because the company is breaking the rules. The company is simply charging a price that allows it to make something like the average rate of profit in the economy.

The phenomenon of increasing returns provides but one more way of explaining why markets fail. If we put all of the explanations of market failure together—inadequate pricing, externalities (spillovers), incomplete contracts, market power, and increasing returns—we arrive at a more complete understanding of the concept of market failure. As noted above, the term refers to any situation in which the market interactions of buyers and sellers result in outcomes that are undesirable either to individuals or to society as a whole. The accompanying box ("The Invisible Foot . . .") provides a list of some of the more common types of market failure.

The type of market failure that stems from increasing returns leads at least some countries to avoid relying on privately owned companies for the supply of such things as electric power generation, transportation networks, and phone systems, all of which are usually characterized by increasing returns. Such countries choose, instead, to have either their governments or regulated enterprises carry on these types of economic activities. Whether these solutions work better in practice than private production without regulation depends on the nature of the governments in question.

The fourth category of market failure listed in the "Invisible Foot" box raises, again, the issue of efficiency and income distribution discussed in Chapter 3. The question is whether one can say that an economy is efficiently allocating a society's resources if some people have huge incomes while others do not have enough income to provide for their most basic needs (see the box "Efficiency, Profitability, and 'Pareto Optimality'" in Chapter 3, p. 63). The general problem is brought out here in the box "Voting With Dollars," and the box on "Sleeping Sickness" provides a concrete illustration of it.

THE INVISIBLE FOOT: WHEN MARKETS FAIL

Market failures occur when the market interactions of buyers and sellers result in outcomes that are undesirable either to individuals or to society as a whole. Thus, market failures result

- *when markets are controlled by a small number of buyers or sellers.* (For example, when there is a single monopoly seller, the price charged for a good will exceed the cost to the firm of producing another unit of the good.)
- *when environmental degradation or other negative externalities resulting from production occur.* (In this case, the cost to the firm of producing another unit of the good will not be the same as the cost to society—the social cost will be greater than the private cost—and the price charged by the firm will not reflect the true scarcity of the product.)
- *when externalities in consumption are present.* (Here, the benefit or cost to the individual consumer will not accurately measure the benefit or cost to society as a whole. An example of a positive consumption externality would be another person's—or a whole neighborhood's—enjoyment of one family's beautiful garden; an example of a negative consumption externality would be the imposition of unwanted smoke by a smoker on nonsmokers.)
- *when people's needs are not reflected in market demands.* (This may happen when individuals, such as homeless people, do not have enough money to purchase necessities, such as housing, for themselves in the market, or it may result from a person's mistaken belief about the ability of a certain good—say, one more gadget—to satisfy his or her needs.)

More recently, some economists have argued that success stories such as London's congestion fee system (see box "Private Incentives, Public Benefits") are the exception, not the rule. They point out that governments generally cannot be trusted—any more than markets can—to perform in accordance with some ideal. Such skeptics have warned that granting a government the right to interfere with the workings of markets may do more harm than good. They have coined the term *government failure* to suggest that the government might not be any more successful than an unfettered market would be in accomplishing a task or solving a problem. According to this view, moreover, giving government more of a role might increase the chances for favoritism, bribery, or other forms of corruption, all of which would result in less-than-optimal allocations of a society's resources.

Still other economists, following the lead of Ronald Coase (see Chapter 4), have favored improving the nature of contracts—basing them on more precisely defined property rights—so that more of the relevant social interactions among economic actors will be

VOTING WITH DOLLARS

It is sometimes said that markets are like elections, in which consumers "vote" with their dollars for the commodities they want. If a large number of dollar "votes" are cast for yellow shirts, a large number of yellow shirts will be produced. In a capitalist economy competition for profits will see to it that resources are allocated in such a way as to produce commodities in the proportions determined by dollar votes.

Voting for commodities in markets is an unusual kind of election, however, because some people vote more times than others. If every dollar of household income had been cast as a vote in 2002, the average household in the richest fifth of the U.S. population would have had more than 14 times as many votes (143,559) as the average household in the poorest fifth (9,931). Rather than the one-person, one-vote principle of democracy, this is more like an economic version of ballot-box stuffing.

Source: U.S. Census Bureau, "Household Income (2002)" available at: http://www.census.gov/hhes/www/income.html.

SLEEPING SICKNESS: "IT REALLY IS A FAILURE OF CAPITALISM."

Sleeping sickness is a horrible disease common in Africa; it attacks the brain, driving the afflicted insane before killing them. Spread by the tsetse fly, it strikes more than a quarter of a million people each year. A cure for the disease has been discovered, "efornithine," and it is so effective that even comatose sleeping sickness patients have been revived. Grateful Africans have dubbed it the "resurrection drug."

Wonderful news. A triumph of modern medicine? Not exactly. Although researchers have known since the early 1990s that efornithine is effective in the treatment of sleeping sickness, it was not put into production because early hopes that the drug would also cure cancer were not borne out. There are no markets (or profits to be made) for something that saves only poor people: they cannot afford to buy it. Now, however, the Bristol Myers Squibb Company is producing efornithine because, as an ingredient in something called Vaniqa, it can be marketed as a cream that removes women's facial hair. Bristol Myers recently promoted Vaniqa with a six-page advertising supplement in *Cosmopolitan* magazine that contained the following text: "If the mustache that prevents you from getting close is yours (not his), it may be time for a beauty about face." The price of a two-month supply of Vaniqa, enough to make a woman's moustache disappear, greatly exceeds the cost of producing enough efornithine to rescue a person from sleeping sickness, madness, and death in Africa.

Bristol Myers is now working with the World Health Organization and such groups as Doctors Without Borders to find a way to make efornithine available in a form that can be used to combat sleeping sickness (rubbing on Vaniqa does not help). But Robert Laverty, a Bristol Myers spokesman, expressed concern about the bottom line: "The question is how this will be funded indefinitely."

The possibility of eradicating sleeping sickness in Africa is a happy accident of the market for facial hair treatments in rich countries. About 3 million people in low- and middle-income countries die of malaria, measles, tetanus, and diarrheal diseases every year, but such deaths will probably not be prevented anytime soon. Only 1 percent of the global market for medicines is in Africa, and the people of the state of Connecticut spend more on health care than do the entire populations of the 38 lowest-income nations of Africa. The profitable markets for pharmaceuticals lie elsewhere: three-quarters of the world's drug sales are in the U.S., Europe, and Japan, where less than one-fifth of the world's people reside and where the incidence of disease is relatively low. Private firms carry on about half of global medical research, but less than 5 percent of that is focused on diseases that are common in poor countries. Of the 1,233 drugs licensed worldwide between 1975 and 1997, only 13 were for tropical diseases.

On his Friday evening PBS show *Now*, Bill Moyers asked Microsoft's Bill Gates, "What does it say to you that 11 million children, roughly, die every year from preventable diseases? What does it say to you that of the 4 million babies who die within their first month, 98 percent are from poor countries? What do those statistics tell you about the world?" Gates replied, "It really is a failure of capitalism. You know, capitalism is this wonderful thing . . . But in this area of diseases of the world at large, it's really let us down." Moyers then countered: "But markets are supposed to deliver goods and services to people," and Gates responded, "and when people have money it does . . . Here what we have is . . . not only don't the people with money have the disease, but they don't *see* the people who have the disease. If we took the world and we just reassorted each neighborhood to be randomly mixed up, then this whole thing could get solved. Because you'd look out your window and you'd say, you know, there's [a] mother over there whose child is dying. You know, let's go help that person." The Bill and Melinda Gates foundation is funding efforts to reduce the incidence of disease in Africa.

Sources: Michael Kremer, "Pharmaceuticals and the Developing World," *Journal of Economic Perspectives,* fall 2002, pp. 67–90; Donald McNeil, "Cosmetic Saves a Cure for Sleeping Sickness," *New York Times,* February 9, 2001, p. A1; *NOW with Bill Moyers,* PBS, May 9, 2003.

PRIVATE INCENTIVES, PUBLIC BENEFITS

Until recently the congested traffic in central London crept along at a snail's pace. Confronted with approaching gridlock, Mayor Ken Livingstone tried a radical solution: charge those operating private cars for the congestion costs they impose on others. In February 2003 the city imposed a fee of $8 a day to operate a car in the central part of London. This fee had to be paid by everyone except central London residents. A high-tech computer system kept track daily of who had paid (among the ways to pay, one could use the text message option on a cell phone). License plate number-recognition devices installed throughout the central part of the city nabbed scofflaws.

The traffic congestion costs (like the costs of pollution) that one imposes on others are examples of external diseconomies, or negative externalities. Livingstone's plan forced drivers to take these costs into account. Economists call this "internalizing the externality." The scheme had the effect of increasing the private marginal cost of driving in the center of London by an amount equal to the fee, thus making the private marginal costs more nearly equal to the social marginal costs. Livingstone's plan could have been lifted straight from the pages of an economics textbook: virtually all economics textbooks, including this one, say that people will change their behavior when the potential costs or benefits of their choices are altered. (See our discussion of *shifts* of supply and demand curves in Chapter 8.)

Livingstone's critics thought that his plan would worsen the traffic problem and hurt businesses in the affected area. To the surprise of many, however, it worked. Traffic flowing into central London was reduced by 20 percent, and delays were cut by almost 30 percent. Average traffic speeds in this previously congested area jumped from 9.5 mph to 20 mph. On an average day in March 2003, about 100,000 motorists were paying the congestion fees. The fees they paid—plus the hefty fines imposed on the dwindling number of people who thought, erroneously, that they could beat the system—generated more than $1 million of revenues each day for the city government. A third of the companies in the affected area said that Livingstone's new policy was helping them; only 5 percent said it was hurting their business. After six weeks of operation, half of all Londoners liked the policy, while only a third of them disliked it. Livingstone's popularity ratings hit an all time high, and London's red double-decker buses are no doubt enjoying a popularity surge, too.

Source: "Ken's Coup," *The Economist*, March 22, 2003, p. 39.

taken into account. They focus specifically on the question of how to narrow the gap between the private and the social costs or benefits borne or received by the decision makers in any given situation.

An example of the Coasian approach is the system under which the government sells or otherwise distributes pollution rights, or "emissions permits," that allow companies to emit a certain quantity of pollutants into the atmosphere, soil, or water during a particular time period. Once such rights have been acquired, they can be exchanged among companies in what is, in effect, a market for pollution rights. In such a market pollution rights can be traded for anything from cash to pollution-absorbing forests. What generally happens, however, is that a company builds a state-of-the-art environment-friendly plant and therefore has a surplus of pollution rights—rights it does not need to use because its new plant emits less pollution than most older plants. This company can then sell its unused pollution rights to another company whose plant exceeds the pollution standard set by the government. Under this system the more a company pollutes the more it will have to pay. Moreover, if the price of a pollution right is set correctly, a company will have to take into account in a precise way the harm it does to others when it decides what technology to use when it is planning, say, the construction of a new plant. With this system, its proponents argue, profit and loss calculations will bring about a desired amount of pollution reduction in the most efficient possible way.

Because millions or billions of dollars as well as life and death matters of public health are affected by policies to correct market failures, the special-interest groups that stand to benefit or lose from their adoption often dominate debates about these policies. An as yet unanswered question is: how can ordinary people build up enough influence in policy-making centers such as Washington, D.C., to equal or outweigh the influence that lobbyists, representing a small number of individuals or corporations, exert on the legislators and others who determine which policies will be adopted—or not adopted—to correct for the various types of market failures discussed in this chapter?

SUGGESTED READINGS

Allen E. Buchanan, *Ethics, Efficiency, and the Market* (Totowa, N.J.: Rowman & Allanheld, 1985).

John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: The Invisible Hand* (New York: Norton, 1989).

Milton Friedman, *Capitalism and Freedom* (Chicago: University of Chicago Press, 1962).

F. A. Hayek, *The Fatal Conceit: The Errors of Socialism* (Chicago: University of Chicago Press, 1988).

David Jenkins, *Market Whys and Human Wherefores: Thinking About Markets, Politics and People* (London and New York: Cassell, 2000).

Robert Kuttner, *Everything for Sale: The Virtues and Limits of Markets* (New York: Knopf, 1997).

Marie Lavigne, *The Economics of Transition: From Socialist Economy to Market Economy,* 2nd ed. (New York: Palgrave, 1999).

Charles E. Lindblom, *The Market System: What It Is, How It Works, and What To Make of It* (New Haven, Conn.: Yale University Press, 2001).

Amartya Sen, *Poverty and Famines: An Essay on Entitlement and Deprivation* (Oxford: Oxford University Press, 1981).

Cass R. Sunstein, *After the Rights Revolution: Reconceiving the Regulatory State* (Cambridge, Mass.: Harvard University Press, 1990).