Field, B. C., & Field, M. K. (2006). Environmental economics: an introduction. McGraw-Hill series (Economics).

Chapter 3

# Benefits and Costs, Supply and Demand

This and the next chapter contain discussions of certain basic tools of **microeconomics.** The objective is to provide enough of an understanding of fundamental concepts so that they can be used later in analyzing environmental impacts and policies. The current chapter is about benefits and costs. The juxtaposition of these two words indicates that we are going to approach things in a **trade-off,** or **balancing,** mode. Economic actions, including environmental actions, have two sides: On the one side they create value and on the other side they encounter costs. Thus, we must have basic concepts that deal with these two parts of the problem. We look first at the question of value, later at costs.

It needs to be mentioned at the very outset that microeconomic theory is **abstract.** This means that it normally proceeds with simplified models that try to capture the essence of a problem without all the details that one observes in the real world. The reason for this is that we want to reveal basic connections and relationships among the important elements of a problem, relationships that are difficult to see if we just observe the surface richness of the real world. There are dangers in this, of course; one can inadvertently overlook details that do have an important impact in reality. For example, in the past many environmental models have been developed without considering the costs of actually enforcing environmental laws. But in the real world, **enforcement costs** are more than a detail; they can have a great impact on the outcomes of environmental regulations. Thus, we need to be careful that our abstractions truly serve to reveal basic connections and do not cover up important dimensions of problems we are trying to understand.

## Willingness to Pay

The value side of the analysis is based on the fundamental notion that individuals have **preferences** for goods and services; given a choice, they can express preferences for one good over another or one bundle of goods over another bundle. How to make visible this abstract notion of preference? We need to simplify the discussion; in a modern economy there are thousands of different goods and services available, so let us focus on just one of them. We now can present the following fundamental concept: The value of this good to a person is what the person is willing and able to sacrifice for it. Sacrifice what? It could be anything, but it makes most sense to talk about sacrificing generalized purchasing power. Thus, the fundamental idea of value is tied to **willingness to pay;** the value of a good to somebody is what that person is **willing to pay** for it.[1]

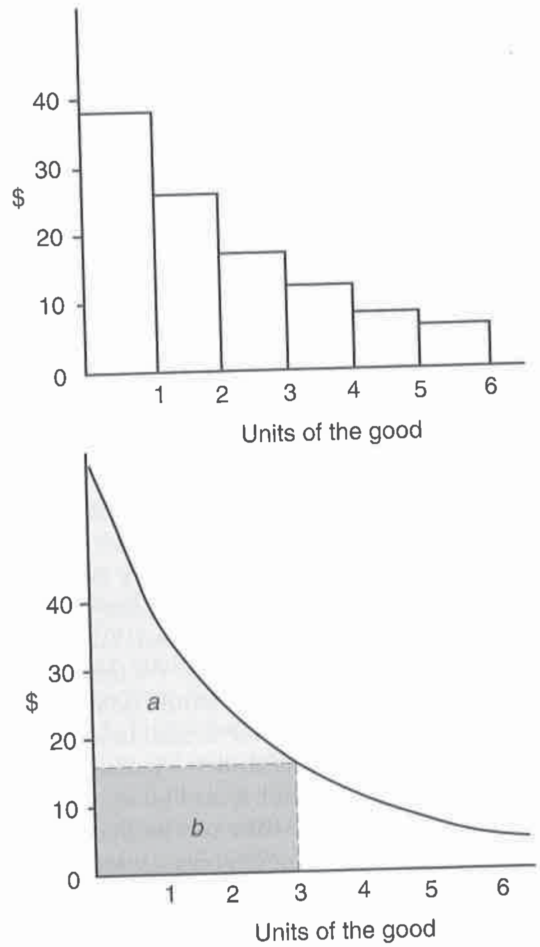
What determines how much a person is willing to pay to obtain some good or service or some environmental asset? It's partly a question of **individual values.** Some people are willing to sacrifice a lot to visit the Grand Canyon; others are not. Some people are willing to pay a lot for a quiet living environment; others are not. Some people place a high value on trying to preserve the habitat of unique animal and plant species; others do not. It is obvious also that a person's **wealth** affects the willingness to sacrifice; the wealthier a person is, the better that person can afford to pay for various goods and services. Willingness to pay, in other words, also reflects **ability to pay.**

Let's consider the willingness to pay of a person for a particular good. We want to build a graphic picture of willingness to pay for various amounts of this good. Assume that the person has none of the good to begin with. We ask her, or perhaps deduce from watching her spend her money, how much she would be willing to pay for a single unit of a good rather than go without. Suppose this is some number, such as $38 pictured in the top of Figure 3.1. We then ask, assuming she already has one unit of this good, how much she would be willing to pay for a second unit. According to Figure 3.1 her answer is $26. In similar fashion, her willingness to pay for each additional unit is shown by the height of the rectangle above that unit: $17 for unit 3, $12 for unit 4, and so on. These numbers depict a fundamental relationship of economics: the notion of diminishing willingness to pay. As the number of units consumed increases, the willingness to pay for additional units of that good normally goes down.

It is not very convenient to work with diagrams that are step-shaped as in the top of Figure 3.1. So we now change things a bit by assuming that people can consume fractions of items in addition to integer values (e.g., as in the number of pounds of bananas consumed per week). What this does is produce a smoothly shaped willingness-to-pay curve, such as the one pictured in the bottom of Figure 3.1. In effect the steps in the willingness-to-pay curve have become too small to see, yielding a smooth curve to work with. On this smooth function we have singled out one quantity for illustrative purposes. It shows that the willingness to pay for the third unit is $17.

[1] It may sound as though we are limiting the analysis only to physical goods and services, but this is not true. The concept of willingness to pay is quite general, and in Chapter 5 we will apply it to differing levels of environmental quality.

**FIGURE 3.1 The Concept of Willingness to Pay**




The next step is to distinguish between total and marginal willingness to pay. Suppose a person is already consuming two units of this good; according to the willingness-to-pay curve, that person would be willing to pay $17 for a third unit. This is the **marginal willingness to pay**—in this case, for the third unit. Marginal is thus a word that describes the *additional* willingness to pay of a person for one more unit. So the height of the rectangles in the top of Figure 3.1 and the height of the curve in the bottom graph show the marginal willingness to pay for this good.

The **total willingness to pay** for a given consumption level refers to the total amount a person would be willing to pay to attain that consumption level rather than go without the good entirely. Suppose the person is consuming at a level of three units; her total willingness to pay for consuming this quantity is $81, which is in fact the sum of the heights of the demand rectangles between the origin and the consumption level in question ($38 for the first plus $26 for the second plus $17 for the third). This corresponds, in the smooth version of the willingness-to-pay function, to the whole area under the willingness-to-pay curve from the origin up to the quantity in question. For three units of consumption, the total willingness to pay is equal to an amount represented by the combined areas *a* and *b*.

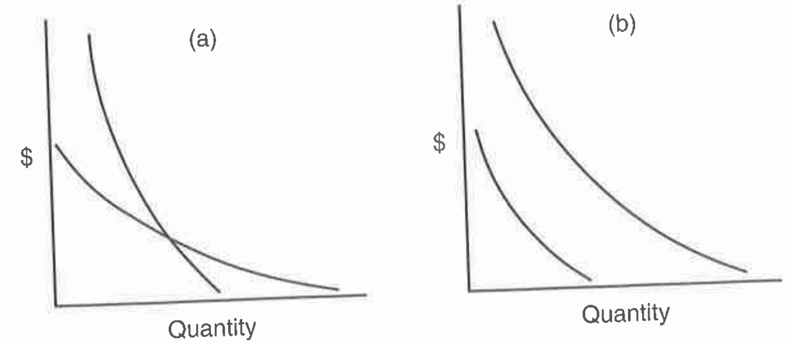
## Demand

There is another way of looking at these marginal willingness-to-pay relationships. They are more familiarly known as **demand curves.** An individual demand curve shows the quantity of a good or service that the individual in question would demand (i.e., purchase and consume) at any particular price. For example, suppose a person whose marginal willingness-to-pay/demand curve is shown in the bottom part of Figure 3.1 is able to purchase this item at a unit price of $17. The quantity he would demand at this price is three units. The reason is that his marginal willingness to pay for each of the first three units exceeds the purchase price. He would not push his consumption higher than this because his marginal willingness to pay for additional quantities would be less than the purchase price.

An individual's demand/marginal willingness-to-pay curve for a good or service is a way of summarizing his personal consumption attitudes and capabilities for that good. Thus, we would normally expect these relationships to differ somewhat among individuals, because individual tastes and preferences vary. Some people are willing to pay more for a given item than other people. Figure 3.2 displays several different demand curves. Panel (a) shows two demand curves, one steeper than the other. The steeper one shows a situation in which marginal willingness to pay drops off fairly rapidly as the quantity consumed increases; while the flatter one shows marginal willingness to pay which, although lower to begin with, goes down less rapidly as quantity increases. These two demand curves could represent the case of one consumer and two different goods or services, or the case of two different consumers and the same good or service.

Panel (b) of Figure 3.2 also has two demand curves; they have the same general shape, but one is situated well to the right of the other. The demand curve lying above and to the right shows a good for which the marginal willingness to pay is substantially higher than it is for the same quantity of the other good. What could account for the difference? They might represent the demand curves of two different people for the same good. But there are other possibilities. How much a person is willing to pay for something obviously depends on how much money she has; more than likely the higher her income, the more she is willing to pay. So the two demand curves in panel (b) could apply to the same individual and the same good, but at two different points in time, the one to the right being her willingness to pay after she has had a substantial increase in her income. The relationship between demand and income is an important one. When the demand for a good or service increases as income increases, we call it

**FIGURE 3.2 Typical Demand/Marginal Willingness-to-Pay Curves**




a **normal good.** Environmental quality is very much a normal good: as their incomes increase, people generally desire higher levels of environmental quality.

There is another way of looking at the two demand curves of panel (b), one that may be very important for the application of these ideas to environmental assets. People's tastes depend on a lot of factors of a psychological and historical kind that are hard to pin down and describe but are nevertheless real. They will depend in part on the experiences that people have and the information they gather over time about the qualities of different goods and how they feel about them. So, for example, the demand curve to the right could be the same consumer's demand curve for a good for which his appreciation has increased over time. For example, these might be his demand curves for outdoor wilderness experiences, the one to the left applying before he knows much about this type of activity and the one to the right applying after he has had some wilderness experiences and learned to like them. Other factors are information and psychology; the demand curve on the right might be a person's demand for a food item before an announcement of the presence of pesticide residues in it, with the curve on the left being the demand curve after the announcement.

Note that the demand curves are in fact curvilinear, rather than straight lines. A straight-line demand relationship would imply a uniform change in the quantity demanded as its price changes. For most goods, however, this is unlikely to be true. At low prices and high rates of consumption, studies have shown that relatively small increases in price will lead to substantial reductions in quantity demanded. At high prices and low quantity demanded, however, price increases have a much smaller effect: they produce much smaller reductions in quantity demanded. This gives us a demand relationship that is convex to the origin (i.e., relatively flat at low prices and steep at higher prices). (See Example 3.1.)

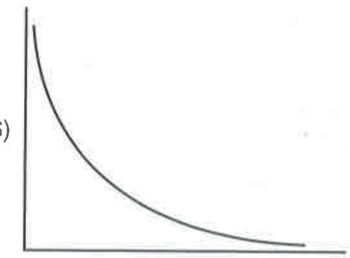
Economics is sometimes misunderstood as assuming that people are driven only by thoughts of their own welfare, that they are complete egoists. Because these are individual demand curves, they do indeed summarize the attitudes of single individuals, but this does not imply that individuals make decisions with only themselves in mind. Some people may indeed act this way, but for most

## The Demand for Water — EXAMPLE 3.1

Researchers have investigated the demand for water by households. Many might think that the amount of water a household uses would be related only to such things as the size of the family rather than the price of the water. This is not the case, however. In general, as the price people pay for water increases, the amount of water they use declines.

This demand is somewhat complicated. Water is used for a number of household purposes—for example, inside the house for sanitation and food preparation and outside the house for car washing, lawn sprinkling, and so forth. At higher prices, consumers will curtail unessential water uses substantially, but their water use for essential purposes will not decline as much in relative terms.

This means that the demand curve for water is shaped as in the diagram.




At low and moderate prices, increased prices will lead to a substantial drop in household water use as people cut back on unessential uses. Thus, the demand curve is relatively flat in this range. But, at higher prices where most of the water is going to essential purposes, further price increases will lead to relatively smaller drops in consumption, hence a steeper demand curve.

there are many other powerful motives that affect their demands for different goods, including altruism toward friends and relatives, feelings of civic virtue toward their communities, a sense of social responsibility toward fellow citizens, and so on. Individual tastes and preferences spring from these factors as well as from more narrow considerations of personal likes and dislikes.

## Aggregate Demand / Willingness to Pay

In examining real-world issues of environmental quality and pollution-control policy, we normally focus our attention on the behavior of groups of people rather than single individuals. Our interest is in the total, or aggregate, demand/marginal willingness to pay of defined groups of people.

An **aggregate demand curve** is the summation of a number of individual demand curves. What individuals are involved depends on which particular aggregation we want to look at: the demand of people living in the city of New York for brussels sprouts; the demand of people living in New Orleans for clean water in the Mississippi River; the demand of people living in the entire country for public parks; and so on. An aggregate demand curve is simply the summation of the demand curves of all the people in the group of interest.

**FIGURE 3.3 Aggregate Demand/Marginal Willingness-to-Pay Curves**

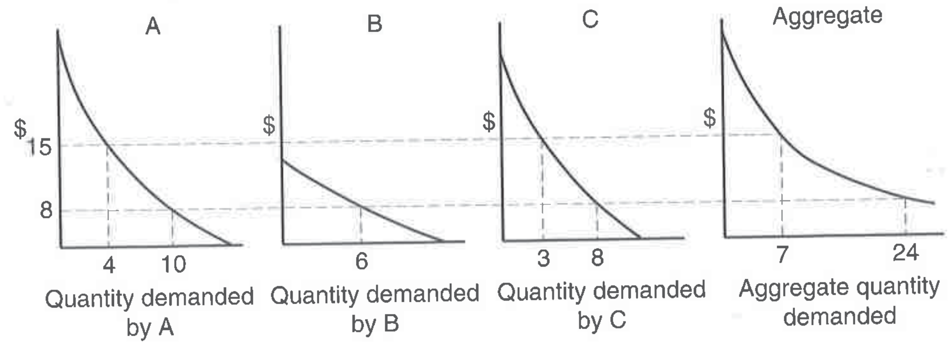


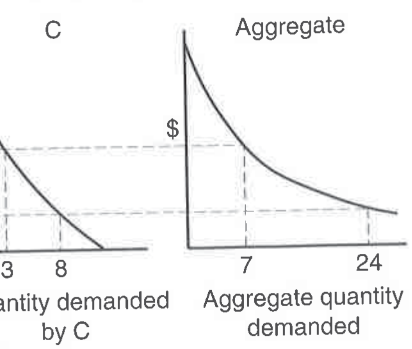

Figure 3.3 depicts a very simple aggregate demand curve, one in which the group consists of only three people. At a price of $8, Person A demands 10 units of this good, whereas at the same price Person B demands 6 units and Person C demands 8 units of the good. Thus, the aggregate demand curve, pictured to the far right, shows an aggregate demand of 24 units for the price of $8. Note that we are summing these individual demand curves horizontally. Looked at in the other direction we note that when Person A is consuming 10 units his marginal willingness to pay is $8, whereas when Persons B and C consume, respectively, at 6 units and 8 units, their marginal willingness to pay is also $8. Therefore, on the aggregate level, the marginal willingness to pay is $8. If one more unit is made available to this aggregate, it must be distributed to Person A, Person B, or Person C, each of whom has a marginal willingness to pay of $8; thus, the aggregate marginal willingness to pay is also $8.

## Benefits

We now come to the idea of **benefits.** Benefit is one of those ordinary words to which economists have given a technical meaning. When the environment is cleaned up, people obtain benefits; when the environment is allowed to deteriorate in quality, benefits are taken away from them—they are, in fact, being damaged. We need some way of conceptualizing and measuring this notion of benefits.

The word *benefits* clearly implies being made better off. If someone is benefited by something, her position is improved—she is better off. Conversely, if she is worse off, it must be because benefits were somehow taken away from her. How do we confer benefits on somebody? We do this by giving him something he values. How do we know that he values something? We know by the fact that he is willing to sacrifice, or willing to pay, for it. According to this logic, then, the benefits that people get from something are equal to the amount they are willing to pay for it.

illingness-to-Pay Curves



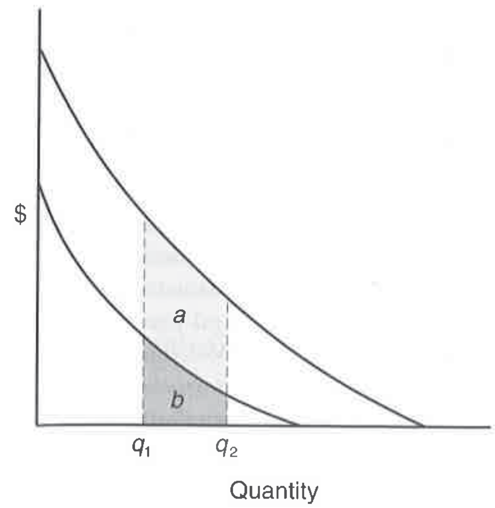

e demand curve, one in which the
e of $8, Person A demands 10 units
on B demands 6 units and Person C
gate demand curve, pictured to the
units for the price of $8. Note that
urves horizontally. Looked at in the
is consuming 10 units his marginal
ons B and C consume, respectively,
ness to pay is also $8. Therefore, on
ss to pay is $8. If one more unit is
distributed to Person A, Person B,
willingness to pay of $8; thus, the
o $8.

fit is one of those ordinary words to
meaning. When the environment is
e environment is allowed to deteri-
rom them—they are, in fact, being
lizing and measuring this notion of

made better off. If someone is bene-
ed—she is better off. Conversely, if
ts were somehow taken away from
dy? We do this by giving him some-
values something? We know by the
to pay, for it. According to this logic,
ething are equal to the amount they



**FIGURE 3.4 Willingness to Pay and Benefits**




The logic behind this definition of *benefits* is quite strong. It means we can use ordinary demand curves to determine the benefits of making various things available to people. For example, Figure 3.4 shows two demand curves, and on the horizontal axis two quantity levels are indicated. Suppose we wish to estimate the total benefits of increasing the availability of this item from quantity $q_1$ to quantity $q_2$. According to our previous thinking, benefits are measured by willingness to pay, and we know that total willingness to pay is measured by areas under the demand curve, in this case the area under the demand curves between quantity $q_1$ and quantity $q_2$. So for the lower demand curve the benefits of such an increase in availability are equal to an amount shown by area *b*, whereas benefits in the case of the higher demand curve are equal to the total area $a + b$.

The logic of this seems reasonable. The people with the higher demand curve must place a greater value on this item; whatever it is, they are willing to pay more for it than the people whose demand curve is the lower function. This is in agreement with common sense. The more people value something, the more they are benefited by having more of that something made available, or, to say the same thing, you can't damage people by taking away from them something that they don't value.

This is the fundamental logic underlying much of environmental economics. It underlies, for example, questions of measuring the damage done to people when the natural environment surrounding them is degraded. It underlies the question of evaluating the impacts of environmental programs and policies undertaken by local, state, and federal governments. This is the strength of the economic approach, the fact that it is based on a clear notion of the value that people place on different things.

But the idea also has shortcomings. For one thing, demand and, therefore, benefits are often very hard to measure when it concerns environmental questions, as we will see in later chapters. For another, we have to remember that demand curves are critically affected by the ability to pay for something as well as preferences. In Figure 3.4, for example, the lower demand curve could represent a group of people with lower incomes than those with the higher demand curve. The logic of the argument would lead to the conclusion that the increase in quantity of $q_2 - q_1$ would produce fewer benefits among lower-income people than among higher-income people. This may not be a very equitable conclusion, depending on the circumstances. Thus, although the logic of the concept is clear, we have to be careful in using it, especially when we are dealing with groups of people with diverse income levels. The main step in doing this is to find out as clearly as possible how the various environmental policies and programs, present or proposed, affect people at different income levels. We discuss this at greater length in later chapters.

One other possible problem exists in using conventional demand curves to measure benefits. An individual's demand for something is clearly affected by how much she knows about it; a person would not be willing to pay for a good if, for example, she was ignorant of its very existence. In Figure 3.4, the higher demand curve might be the demand for a good before it is found out that it contains a carcinogenic substance, and the lower demand curve shows demand after this fact becomes known. There is nothing especially surprising about this; people after all do become more knowledgeable about things over time as a matter of course. But in today's world this could be a complication, especially with regard to the environment. We don't fully understand many of the effects of environmental degradation; furthermore, people's views about the importance of many of these effects are blown back and forth almost from day to day, by the media, by the scientific press, and so on. Care must be exercised in taking people's demand curves of the moment, influenced as they are by all kinds of real and imagined factors, as true expressions of the benefits of environmental actions. It is not that they are irrelevant; it is only that they have to be taken with a certain amount of caution.

## Cost

We now switch to the other side of the picture and consider costs. Although some things in life are free—an idea, for example—it is generally true that goods and services cannot be produced out of thin air; they require the expenditure of productive resources, or inputs, in the process. The more of something that is desired, the more resources we will have to devote to its production. What is needed is a way of describing and talking about the costs of producing useful things, whether these are normal consumer goods, such as cars or hot-water bottles, or services, such as transportation or insurance, or environmental quality through the treatment of waste residuals, recycling, or land-use controls.

Imagine a simple production process. Suppose, for example, we are producing a certain line of cardboard boxes. To produce boxes, many types of productive inputs are required: labor, machinery of various descriptions, energy, raw materials, waste-handling equipment, and so on. The first thing needed is a way of valuing these productive resources. If we are a private firm operating in a market economy, we would have little problem: We would value them according to what they cost to procure in the markets. Our profit-and-loss statement at the end of the year would reflect the monetary out-of-pocket costs of the inputs used in the production operation. But our concept of cost will be broader than this. From this wider perspective the costs of these cardboard boxes are what could have been produced with these productive inputs had they not been used in box production. The name for this is **opportunity cost.**

## Opportunity Cost

The **opportunity cost** of producing something consists of the maximum value of other outputs we could and would have produced had we not used the resources to produce the item in question. The word *maximum* is used for a reason. The productive inputs used to produce the cardboard boxes could have been used to produce a variety of other things, perhaps automobiles, books, or pollution-control equipment. The opportunity cost of the boxes consists of the maximum value of the alternative output that could have been obtained had we used these resources differently.

Opportunity costs include out-of-pocket costs but are wider than this. Some inputs that are actually used in production may not get registered as cash costs. For example, the spouse of the cardboard box plant operator works as an unpaid assistant in the front office. This may not register as an out-of-pocket cost, but he certainly has an opportunity cost because he could have been working somewhere else if he was not working here. Even more importantly for our purposes, the cardboard box manufacturing process may produce waste products that are pumped into a nearby stream. Downstream these production residuals produce environmental damage, which are real opportunity costs of producing cardboard boxes, even though they do not show up as costs in the plant's profit-and-loss statement.

The opportunity cost idea is relevant in any situation in which a decision must be made about using productive resources for one purpose rather than another. For a public agency with a given budget, the opportunity costs of a particular policy are the value of alternative policies it might have pursued. For a consumer, the opportunity cost of spending time searching for a particular item is the value of the next most valuable thing to which the consumer could have devoted time.

How is opportunity cost measured? It is not very useful to measure it in terms of the number of other physical items that could have been produced. Nor is there enough information in most cases to be able to measure the value of the next best output that was forgone. In practice, therefore, opportunity costs are measured by the market value of inputs used up in production. For this to work, we have to take care that the inputs have been correctly valued.

The office labor must be valued at the going rate even though it is not paid in practice. The effects on downstream water quality must be evaluated and included. Once all inputs have been accounted for, their total value may be taken as the true opportunity costs of production.

## Private and Social Costs

Another important distinction is that between **private costs** and **social costs.** The private costs of an action are the costs experienced by the party making the decisions leading to that action. The social costs of an action are *all* of the costs of the action, no matter who experiences them. Social costs include private costs, but also may include much more in certain situations.

Consider the action of driving a car. The private costs of this include the fuel and oil, maintenance, depreciation, and even the driving time experienced by the operator of the car. The social costs include all these private costs and also the costs experienced by people other than the operator who are exposed to the congestion and air pollution resulting from use of the car. This distinction between private and social costs will be very important in later sections where we begin to analyze environmental problems with these tools.

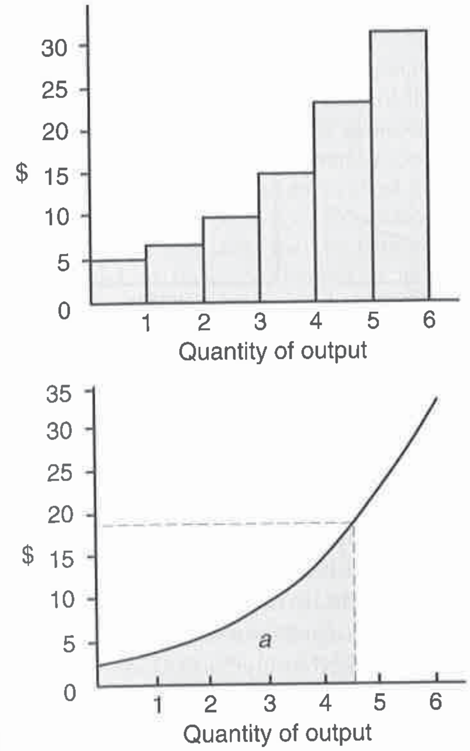
## Cost Curves

To summarize cost information, we use cost curves, which are geometric representations of the costs of producing something. And, just as in the case of willingness to pay, we differentiate between **marginal costs** and **total costs.** Consider the cost curves in Figure 3.5. They are meant to apply to a single producing organization, a firm, or perhaps a public agency that is producing some good or service. The graph is laid out, the same as in previous graphs, with quantity on the horizontal axis and a monetary index on the vertical axis. The quantity relates to some period of time, such as a year. The top panel shows marginal costs in terms of a step-shaped relationship. It shows that it costs $5 to produce the first unit of output. If the firm wants to increase output to two units, it must spend an added $7. The addition of a third unit would add $10 to total costs, and so on. Marginal cost is a symmetrical measure; it is the added costs, the amount by which total costs increase, when output is increased by one unit. It is also the cost savings if production were to decrease by one unit. Thus, the reduction in output from four to three units would reduce total costs by $15, the marginal cost of the fourth unit.

It is inconvenient to work with step-shaped curves, so we make the assumption that the firm can produce intermediate quantities as well as integer values. This gives a smooth marginal cost curve, as shown in the bottom panel of Figure 3.5. This curve now shows the marginal cost—the added cost of one more unit of output—for any level of output. For example, at an output level of 4.5 units, marginal cost is $19.

Marginal cost curves can be used to determine **total production costs.** On the stepped marginal cost curve of Figure 3.5, suppose we want to know the total cost of producing five units of this item. This is equal to the cost of the first unit ($5), plus that of the second ($7), plus that of the third ($10), and so on.

**FIGURE 3.5** **The Concept of Marginal Cost**




This total is $60; geometrically this is equal to the total area of the rectangles above the first five units of output. Analogously, in the smoothly shaped marginal cost function in the bottom of the diagram, the total cost of producing a given quantity is the dollar amount equal to the area under the marginal cost curve between the origin and the quantity in question. The total cost of producing 4.5 units of output is thus given by the area marked *a* in the figure.

## The Shapes of Cost Curves

The height and shape of the marginal cost curve for any production process will differ from one situation to another, based on several underlying factors. A key determining factor is the technology utilized in production, and we discuss this concept later. The price of inputs is also an important factor influencing the heights of marginal cost curves. In general, if input prices increase to a firm or group of firms, their marginal cost curves will shift upward. Another important element is *time*, specifically the amount of time that a firm has to adjust to changes in its rate of output. These factors may be better understood by looking at some actual marginal cost curves.

**FIGURE 3.6 Typical Marginal Cost Curves**

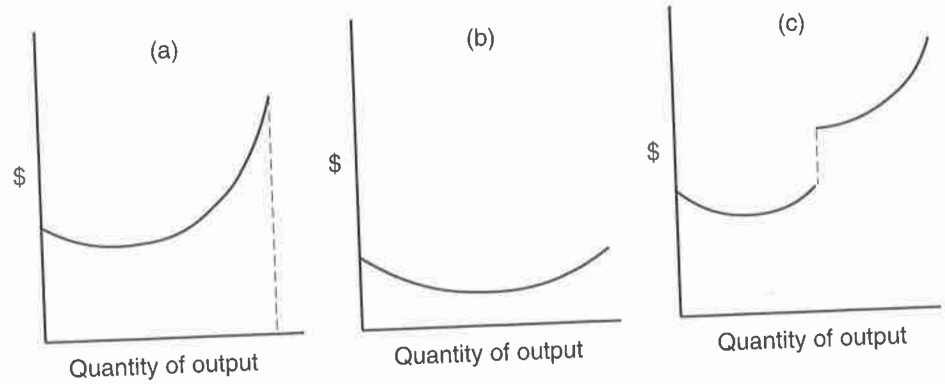
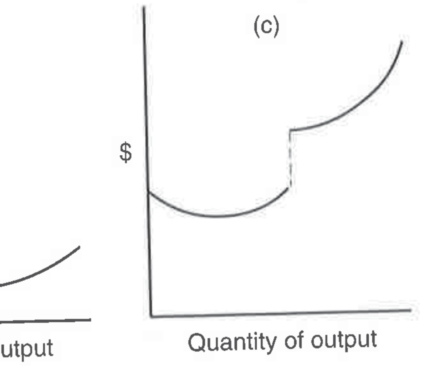



Figure 3.6 shows several marginal cost curves. Panel (a) shows a very typical marginal cost curve; initially it declines as output increases but then it increases as output gets larger. The initial decline comes about because of basic efficiencies achievable with larger quantities at this level. Suppose our "output" refers to the quantity of wastewater handled in a municipal treatment plant. At very low levels of output, the plant is not being fully utilized; thus, output increases in this range are accompanied by less than proportionate increases in production cost, giving marginal costs that diminish. But as output increases, the capacity of the plant is approached. Machinery must be worked longer, additional people must be hired, and so on. Thus, marginal cost begins to increase. As the capacity of the operation is neared, these problems become more acute. To continue to increase output, more extraordinary measures are required, which can only be done at a high cost; thus, marginal cost increases even more. A point may come at which it becomes almost impossible to increase output further, which is the same as saying that the marginal costs of production at this point increase without limit. This limit is indicated by the vertical dashed line in panel (a) of Figure 3.6.

This marginal cost curve depicts an important generic characteristic of all marginal cost curves, namely, that although they may initially decline, they will always increase, eventually, as output becomes large enough. These increases are related to certain underlying factors, such as increased plant utilization, the need to reach farther away for raw materials, and the inevitable higher management costs that accompany larger operations. Virtually all economic studies of particular operations and industries demonstrate increasing marginal production costs, and this fact will be an important shaping element in our later discussions specifically related to environmental quality management. (See Example 3.2.)

Panel (b) of Figure 3.6 shows a marginal cost curve similar in general shape to the one in panel (a), but with less pronounced curvature. In particular, although this marginal cost curve eventually increases, it does so less steeply than the first one. This is more typical of a long-run marginal cost curve, that is,




urves. Panel (a) shows a very typical
utput increases but then it increases
mes about because of basic efficien-
s level. Suppose our "output" refers
municipal treatment plant. At very
fully utilized; thus, output increases
roportionate increases in production
But as output increases, the capacity
be worked longer, additional people
st begins to increase. As the capacity
become more acute. To continue to
sures are required, which can only
t increases even more. A point may
le to increase output further, which is
of production at this point increase
e vertical dashed line in panel (a) of

nportant generic characteristic of all
h they may initially decline, they will
mes large enough. These increases are
s increased plant utilization, the need
nd the inevitable higher management
rtually all economic studies of partic-
rate increasing marginal production
aping element in our later discussions
ty management. (See Example 3.2.)
al cost curve similar in general shape
ronounced curvature. In particular,
ally increases, it does so less steeply
long-run marginal cost curve, that is,

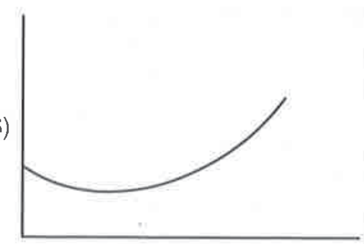


## The Marginal Costs of Producing Pizzas

**EXAMPLE 3.2**

Suppose a local pizza delivery operation has a baking facility and a fleet of three cars and plans to devote their entire operation to making and delivering pizzas. Let us consider the marginal cost of pizza production and delivery. The marginal cost of the first pizza may be fairly high because to begin production at all requires a certain minimum set of inputs. At slightly higher, but still low, rates of production, added output might be obtainable with relatively modest increases in cost because the oven is not being used intensively and a little added flour, sauce, spices, and gas for the cars may be all that is needed. In other words, at low levels of production, we might expect the marginal cost of pizza production to be relatively low or even declining. But at somewhat higher levels of production, marginal cost can be expected to increase. Once the oven is used intensively, larger amounts of other inputs (ingredients, more preparation space, added labor, more car repair, etc.) would be necessary to push production higher. At even greater levels of output, we would expect marginal costs to increase sharply as the overall limits of the facility in terms of pizza production and delivery are reached. All these considerations might be expected to give us a marginal pizza production cost function something like the one in this example.

All the reasoning of the preceding paragraph was based on the assumption that the owner had one oven and three cars. This is what economists call a **short-run** situation, because one or more of the essential production inputs is fixed in amount. In the longer run, of course, the owner could obtain more ovens and cars and therefore be able to obtain increases in output at marginal costs lower than those pictured in the diagram. Yet even with increases in facilities we would expect marginal costs eventually to increase because it will become more costly to bid additional equipment away from other uses and because it will become more difficult to coordinate and carry out decisions over an ever-increasing size of operation. Thus, even in the **long run**, when all production inputs are freely variable in quantity, we would expect marginal costs to increase. Eventually, increasing marginal production costs characterize not just pizza production but the production of most goods and services, with appropriate differences of course in the technologies with which production is pursued in different circumstances.




one where enough time is given for operators of firms to adapt fully to an increase in the rate of output. In the short run, our wastewater treatment plant had a certain capacity that was basically fixed; but in the long run, there is time to build a larger treatment plant with higher capacities. For larger outputs, the marginal costs of this larger plant will be lower than those of the smaller plant.

Yet, even in these long-run situations marginal costs will eventually increase, as is depicted in panel (b). *In our subsequent discussions we will assume that we are working with long-run marginal cost curves, unless specified otherwise.*

Panel (c) of Figure 3.6 represents a more complicated case where there is a discontinuity in the marginal cost curve. After a short downward section, the marginal costs generally trend upward, and at one point they jump upward by some amount. This might represent a "lumpy" investment in new types of technology at a certain point as output increases.

## Technology

The most important factor affecting the shapes of marginal cost functions is the **technology** of the production process. By technology we mean the inherent productive capabilities of the methods and machines being employed. Any modern production requires capital goods (machinery and equipment) of various types and capacities, labor inputs, operating procedures, raw materials, and so on. The quantity of output a firm can get from a given set of inputs depends on the technical and human capabilities inherent in these inputs. The marginal cost curves pictured in Figure 3.6 could relate to different industries because the marginal cost curves are so different. But even within the same industry marginal cost curves can differ among firms. Some firms will be older than others; they may be working with older equipment that has different cost characteristics. Even firms of the same age may have different production techniques; past managerial decisions may have put them in different positions in terms of the marginal production costs they experience today.

This concept of technology is vitally important in environmental economics because **technological change** can provide ways to produce goods and services with fewer environmental side effects and also better ways of handling the quantities of production residuals that remain. In our simple cost model, technical advancement has the effect of shifting marginal cost curves downward. Technological progress makes it possible to produce a given increase in output at a lower marginal cost. It also reduces total production cost. Consider Figure 3.7. $MC_1$ is the firm's marginal cost curve before a technical improvement; $MC_2$ is the marginal cost curve after some technical improvement has been put into effect. The technical change, in other words, shifts the marginal cost curve downward. We also can determine how much total production costs are reduced as a result of the technological change. Consider output level $q^*$. With $MC_1$ the total annual cost of producing output $q^*$ is represented by the area $a + b$, whereas after the reduction in the marginal cost curve to $MC_2$ the total annual cost of producing $q^*$ is equal to area $b$. Thus, the reduction in total cost made possible by the technological change is equal to area $a$.

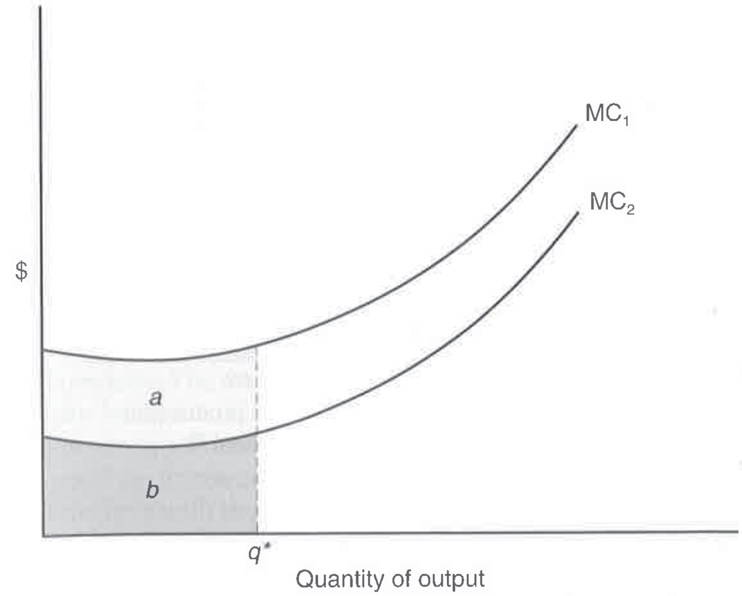
Technological change does not normally happen without effort; it normally requires research and development (R&D). R&D in environmental industries is obviously an important activity to promote, and one of the criteria we will want to use to evaluate environmental policies is whether the policies create incentives

costs will eventually increase, as
*ussions we will assume that we are*
*s specified otherwise.*
omplicated case where there is a
r a short downward section, the
one point they jump upward by
investment in new types of tech-

s of marginal cost functions is the
nology we mean the inherent pro-
nines being employed. Any mod-
inery and equipment) of various
procedures, raw materials, and
rom a given set of inputs depends
rent in these inputs. The marginal
to different industries because the
en within the same industry mar-
ne firms will be older than others;
that has different cost characteris-
ferent production techniques; past
different positions in terms of the
oday.
ortant in environmental economics
ays to produce goods and services
also better ways of handling the
in. In our simple cost model, tech-
marginal cost curves downward.
produce a given increase in output
production cost. Consider Figure 3.7.
ore a technical improvement; $MC_2$
cal improvement has been put into
ds, shifts the marginal cost curve
h total production costs are reduced
sider output level $q^*$. With $MC_1$ the
is represented by the area $a + b$,
l cost curve to $MC_2$ the total annual
us, the reduction in total cost made
al to area $a$.
happen without effort; it normally
R&D in environmental industries is
, and one of the criteria we will want
whether the policies create incentives



**FIGURE 3.7** **Technological Improvement**

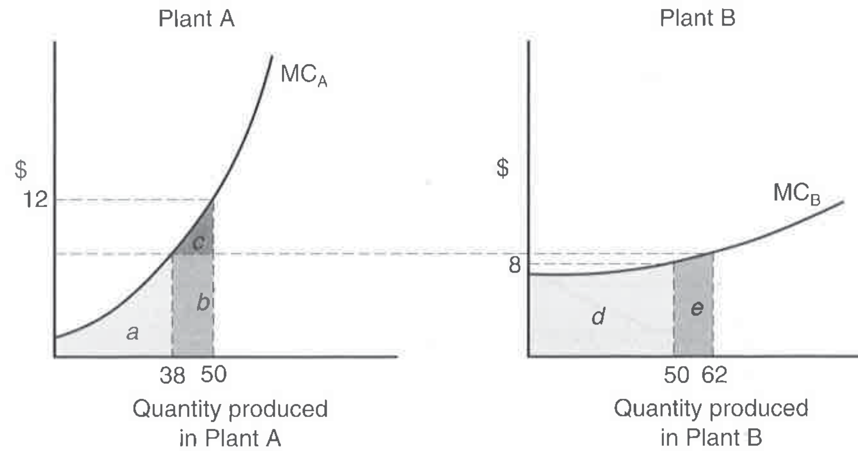



for individuals, firms, and industries to engage in vigorous R&D programs. In very simple terms, the incentive to engage in R&D is the cost savings that result from the new techniques, materials, procedures, and so on, that are discovered in the effort. The cost savings shown in Figure 3.7 (area $a$) show part of this incentive. This is the cost savings that would result each year, and the accumulation of these annual cost savings represents the full R&D incentive.

## The Equimarginal Principle

We come now to the discussion of a simple but important economic principle, one that is used repeatedly in chapters to come. It is called the **equimarginal principle.** To understand it, take the case of a firm producing a certain product and assume that the firm's operation is divided between two different plants. For example, suppose there is a single power company that owns two different generating plants. Each plant produces the same item, so that the total output of the firm is the sum of what it produces in the two plants. Assume that the plants were built at different times and make use of different technology. The old one, Plant A in Figure 3.8, has older technology; it has a marginal cost curve that starts relatively low but rises steeply as production increases. The new plant, Plant B in Figure 3.8, uses newer technology; it has a higher marginal cost at low output levels, but marginal costs do not rise as steeply as production increases.

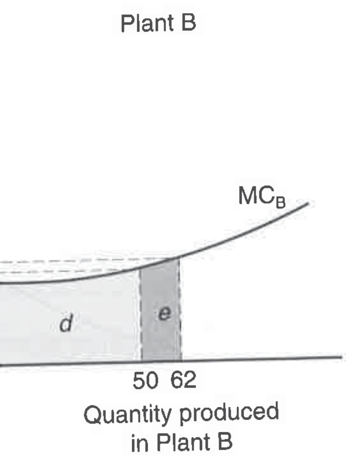
**FIGURE 3.8** **The Equimarginal Principle**




Consider now a situation in which this two-plant firm wants to produce a total output of, say, 100 units. How many units should it produce in each plant in order to produce the 100 units at the *least total cost?* Would it be best to produce 50 units in each plant? This is depicted in Figure 3.8; at an output of 50, Plant A has a marginal cost of \$12 whereas Plant B has a marginal cost of \$8. Total production costs are the sum of total costs at each plant, or $(a + b + c) + d$. Here is the important point: The total cost of the 100 units can be lowered by reallocating production. Reduce production in Plant A by one unit and costs will fall by \$12. Then increase the production in Plant B by one unit and costs there will rise by \$8. Total output is still 100 units, but there has been a cost saving of \$12 − \$8 = \$4. Thus, total cost, the sum of the costs in the two plants, has gone down.

As long as the marginal costs in the two plants differ from one another, we can continue to reallocate production—away from the high-marginal-cost plant and toward the low-marginal-cost plant—and get a reduction in total cost. In fact, the total costs of producing the 100 units in the two plants will be at a minimum only when the marginal costs of the two plants are equal, hence the "equimarginal principle." In the figure, this happens when the output in Plant A is 38 units and the output in Plant B is 62 units. Total costs in geometric terms are now $a + (d + e)$.

The equimarginal principle therefore says the following: If you have multiple sources to produce a given product or achieve a given goal, and you want to minimize the total cost of producing a given quantity of that output, distribute production in such a way as to equalize the marginal costs between the production sources. There is another way of saying it that may look different but actually is not: If you have a given amount of resources and you want to maximize the total amount produced, distribute total production among the




two-plant firm wants to produce a
nits should it produce in each plant
*t total cost?* Would it be best to pro-
ed in Figure 3.8; at an output of 50,
s Plant B has a marginal cost of \$8.
osts at each plant, or $(a + b + c) + d$.
of the 100 units can be lowered by
n in Plant A by one unit and costs
action in Plant B by one unit and
ll 100 units, but there has been a cost
e sum of the costs in the two plants,

o plants differ from one another, we
ay from the high-marginal-cost plant
and get a reduction in total cost. In
its in the two plants will be at a min-
he two plants are equal, hence the
s happens when the output in Plant A
units. Total costs in geometric terms

ys the following: If you have multiple
hieve a given goal, and you want to
en quantity of that output, distribute
the marginal costs between the pro-
of saying it that may look different
mount of resources and you want to
stribute total production among the



sources in such a way as to equalize marginal costs. This principle will be very valuable when we take up the issue of getting maximum emissions reductions from given amounts of resources.

## Marginal Cost and Supply

A critical question in the analysis of any economic system is whether private profit-seeking firms (as well as public, politically minded agencies) will produce the correct quantities of output from the standpoint of society as a whole, not only for conventional items such as cardboard boxes, but also for less conventional items such as the amounts of environmental quality. To address this question one must understand how firms normally determine the quantities they will produce. The marginal cost of production is a key factor in determining the **supply** behavior of firms in competitive circumstances. In fact, the marginal cost curve of a firm acts essentially as a **supply curve,** showing the quantity of the good the firm would supply at different prices. Consider Figure 3.9. Assume that the firm with the indicated marginal cost curve is able to sell its output at a price of $p^*$. The firm will maximize its profits by producing that quantity of output where marginal cost is equal to $p^*$; that level is designated $q^*$. At any output

**FIGURE 3.9** **Marginal Cost and Supply**

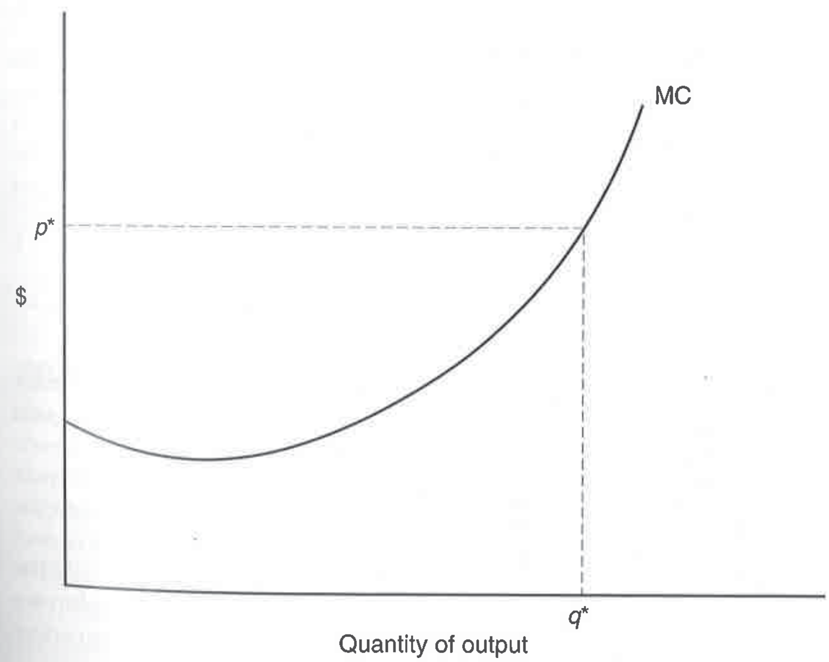
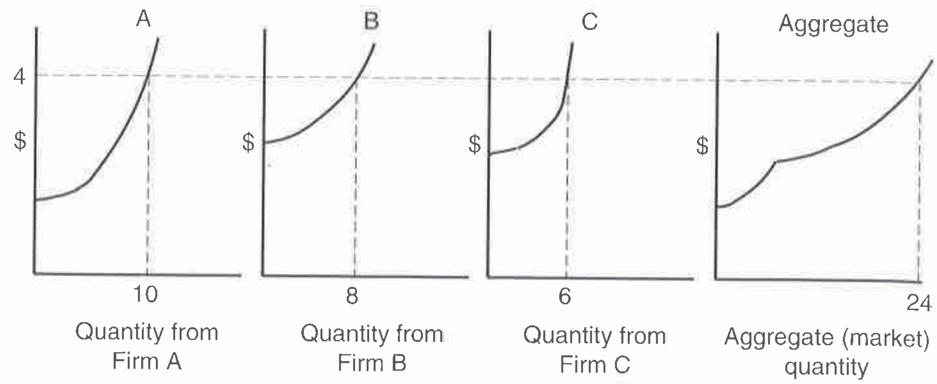


**FIGURE 3.10** **Derivation of Aggregate (Market) Supply from Individual Firm Supply Curves**




level less than this, MC < $p^*$, so a firm could increase its profits by increasing output. At any output level above this, $p^*$ < MC, so a firm is actually producing items for which the marginal cost is higher than price; in this case, the firm should reduce output if it wishes to maximize its profits.

We are often interested in the supply performance of industries composed of many firms rather than that of individual firms. An **aggregate supply** curve shows the amounts supplied by a collection of firms all producing the same output. The idea is analogous to the concept of aggregate demand we had in the previous section. The aggregate supply curve of a group of firms is the sum of the individual supply curves of all the firms in the group. This is depicted in Figure 3.10. There are three firms, A, B, and C, with marginal cost curves as depicted in the first three panels of the figure. At a common price, say $4, Firm A supplies 10 units, Firm B supplies 8 units, and Firm C supplies 6 units. Thus, the aggregate supply at that price is 24 units, as depicted in the far right panel of Figure 3.10.

## Summary

In this chapter we covered briefly some of the basic tools of microeconomics. Later chapters will rely heavily on these ideas, especially on the equimarginal principle and on graphs, where we will want to jump back and forth between marginal and total measures. When we begin to look at real-world problems of environmental analysis and policy design, it is easy to get pulled so far into the countless details that basic economic ideas get lost. It is the fundamental economic building blocks, such as those in this chapter, that allow us to identify the primary economic characteristics of these problems and proceed to develop solutions to them.

supply, willingness to pay,
ory microeconomic texts. The
at deeper explanation of these
o consult the appropriate
ore popular texts are the

*omics: Explore and Apply,*
e River, NJ, 2005, Chapters 6, 7.

*economics, Principles and Policy,*
X, 2000, Chapters 5, 6, 8.

*nomics*, 6th ed., Houghton

*croeconomics*, 7th ed., Prentice
4, 6, 7.

*s of Economics*, 2nd ed.,
ers 4, 5, 6, 7.

istopher T. S. Ragan:
gman, Reading, MA, 1999,

3rd ed., South-Western,

e: *Economics*, 16th ed.,
ers 20, 21, 22.

lison Wesley, Reading, MA, 2005,

us: *Economics*, 18th ed.,
ers 6, 8.

10th ed., McGraw-Hill/Irwin,

# Chapter 4

# Economic Efficiency and Markets

This chapter has several objectives. First is to develop the notion of **economic efficiency** as an index for examining how an economy functions and as a criterion for judging whether it is performing as well as it might. Economic efficiency is a simple idea but one that has much to recommend it as a criterion for evaluating the performance of an economic system or a part of that system, but it has to be used with care. A single firm or group of firms may be judged very efficient in their own limited way as long as they are keeping costs low and making a profit. Yet, to evaluate the *social* performance of these firms, we must use the idea of economic efficiency in a wider sense. In this case it must include all the social values and consequences of economic decisions—in particular, environmental consequences. It is important also to discuss the relationship between economic efficiency and **economic equity.**

The second task is to address the question of whether a **market system,** left to itself, can produce results that are socially efficient. We will see that there are cases in which a system of private markets will not normally be able to bring about results that are efficient in this wider sense. This leads into the next chapter, where we will examine the policy question; that is, if the economy is not operating the way we want it to, especially in matters of environmental quality, what kind of public policy might be used to correct the situation?

Economic efficiency is a criterion that can be applied at several levels: to input usage and to the determination of output levels. We are going to concentrate in this chapter on the second of these because ultimately we want to apply the concept to the "output" of environmental quality. There are two questions of interest: (1) What quantity ought to be produced and (2) what quantity is produced in fact? The first question deals with the notion of efficiency, the second with the way markets normally function.

## Economic Efficiency

In the preceding chapter we introduced two relationships, that between the quantity of output and willingness to pay, and that between output and marginal production costs. Neither of these two relationships, by itself, can tell us what the most desirable level of output is from society's standpoint. To identify this output level, it is necessary to bring these two elements together. The central idea of **economic efficiency** is that there should be a balance between the value of what is produced and the value of what is used up to produce it. In our terminology, there should be a balance between willingness to pay and the marginal costs of production.

Efficiency is a notion that has to have a reference point. It is critical to ask: efficient from the standpoint of whom? What is efficient for one person, in the sense of balancing costs and benefits, may not be efficient for somebody else. We want to have a concept of efficiency that is applicable to the economy as a whole. This means that when referring to marginal costs, *all* the costs of producing the particular item in question must be included, no matter to whom they accrue. When dealing with marginal willingness to pay, we must insist that this represents accurately *all* of the value that people in the society place on the item. This does not necessarily mean that all people will place the same value on all goods; it means only that there are no missing sources of value.

How do we identify the rate of output that is socially efficient? Suppose we focus on a particular type of output; in practice it could be refrigerators, automobiles, college educations, or a certain type of pollution-control equipment. Suppose that our item is currently being produced at a particular rate, and we wish to know whether it would benefit society to have this output level increased by a small amount. To answer this requires comparing the marginal willingness to pay for that extra output with the marginal opportunity costs of the output. If the former exceeds the latter, we would presumably want the extra output to be produced; otherwise, we would not.

This can be analyzed graphically by bringing together the two relationships discussed in the last chapter. Figure 4.1 shows the aggregate marginal willingness-to-pay curve (labeled MWTP) and the aggregate marginal cost curve (MC) for the good in question. The efficient level of production for this item is the quantity identified by the intersection of the two curves, labeled $q^e$ in the figure. At this output level the costs of producing one more unit of this good are just exactly equal to the marginal value of it, as expressed by the marginal willingness-to-pay curve. This common value is $p^e$.

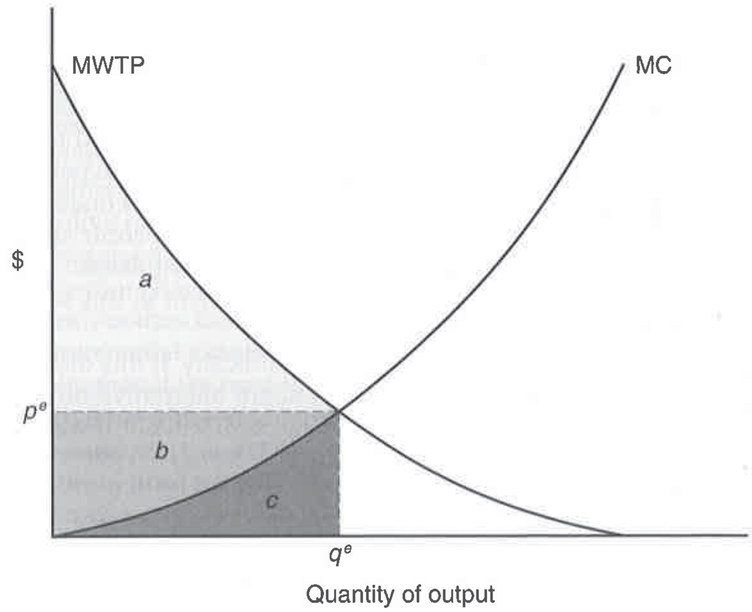
The equality of marginal willingness to pay and marginal production cost is the test for determining if output is at the socially efficient level. There is another way of looking at this notion of efficiency. When a rate of output is at the socially efficient level, the net value, defined as **total willingness to pay** minus **total costs,** is as large as possible. In fact, we can measure this net value on the diagram. At $q^e$ we know that the total willingness to pay is equal to an amount corresponding to the area under the marginal willingness-to-pay curve from

relationships, that between the
l that between output and mar-
lationships, by itself, can tell us
society's standpoint. To identify
two elements together. The cen-
hould be a balance between the
at is used up to produce it. In our
n willingness to pay and the mar-

erence point. It is critical to ask:
is efficient for one person, in the
t be efficient for somebody else.
s applicable to the economy as a
rginal costs, *all* the costs of pro-
be included, no matter to whom
ngness to pay, we must insist that
people in the society place on the
people will place the same value
issing sources of value.
is socially efficient? Suppose we
ce it could be refrigerators, auto-
of pollution-control equipment.
luced at a particular rate, and we
ety to have this output level in-
requires comparing the marginal
the marginal opportunity costs of
we would presumably want the
ould not.
ng together the two relationships
shows the aggregate marginal
and the aggregate marginal cost
icient level of production for this
ion of the two curves, labeled $q^e$ in
ducing one more unit of this good
f it, as expressed by the marginal
e is $p^e$.
y and marginal production cost is
ocially efficient level. There is an-
cy. When a rate of output is at the
as **total willingness to pay** minus
can measure this net value on the
gness to pay is equal to an amount
nal willingness-to-pay curve from



**FIGURE 4.1** **The Socially Efficient Rate of Output**




the origin up to $q^e$; this area consists of the sum of the three subareas: $a + b + c$. Total cost, however, consists of the area under the marginal cost curve, or area $c$. Thus, the surplus is $(a + b + c) - c = a + b$, which is the triangular area enclosed by the marginal willingness-to-pay curve and the marginal cost curve. At any other quantity the corresponding value of total willingness to pay minus total production costs will be less than this area $a + b$.

Let's be clear on what this graph is saying. We noted previously that the marginal willingness-to-pay curve is assumed to represent accurately all the benefits that people in our economy actually experience when the good becomes available. The marginal production cost curve is assumed to contain all the true opportunity costs that are required to produce this good—no hidden or overlooked costs have been left out. Thus, the quantity $q^e$ is **efficient** because it produces a balance between the two sides—between the marginal worth of a good, as indicated by consumers' willingness to pay for it, and what it costs society to produce it, as measured by marginal costs.[1]

[1] The graphs discussed in this and the preceding chapter show the production and consumption of some good or service that has positive value. In later chapters we will adapt them to explore the production of what we might call a "bad," namely, environmental pollution. Then the units along the horizontal axis would be quantities of some pollutant. The marginal cost curve would show the increasing costs, or damages, to society from increasing quantities of pollutants. The demand curve, on the other hand, would show the diminishing marginal savings to polluting firms from being able to emit more pollution into the environment. We will discuss this in much greater detail in Chapter 5.

## Efficiency and Equity

From the standpoint of society at large, production is at an efficient level when marginal benefits equal marginal production costs, that is, when net benefits are maximized *no matter to whom those net benefits accrue.* Efficiency doesn't distinguish among people. A dollar of net benefits to one person is considered to be worth a dollar to anybody else. One hundred dollars of benefits to one person is considered to be worth the same as one dollar of benefits to each of one hundred people. In the real world, an outcome that benefits very rich people at the expense of poor people would be regarded by most people as inequitable. This is simply another way of saying that an outcome that is efficient in this sense need not necessarily be equitable.

**Equity** is tied closely to the distribution of wealth in a society. If this distribution is regarded as essentially fair, then judgments about alternative output levels may justifiably be made using only the efficiency criterion. But if wealth is distributed unfairly, the efficiency criterion by itself may be too narrow. Having said this, however, we have to recognize that in the assessment of economic outcomes, the relative emphasis to be put on efficiency and equity is a matter of controversy. It is controversial in the political arena; it is controversial among economists themselves.

We will have much to say about distributional issues and equity throughout this book. Chapter 6 contains terminology for describing the distributional impacts of environmental policies. Chapter 9 contains a discussion of the role of economic equity as a criterion for evaluating environmental policies.

## Markets

Having specified what economic efficiency means, we next ask whether a market system, a system in which the major economic decisions about how much to produce are made by the more or less unhindered interaction of buyers and sellers, gives results that are socially efficient. In other words, if we rely entirely on the market to determine how much of this item gets produced, will it settle on $q^e$?

Why worry about this question? Why not simply jump to the question of public policy? Doesn't this question imply that, at bottom, we are committed to the market system, and isn't this the system that, from an environmental point of view, has gotten us into trouble in the first place? If the market doesn't do the job, maybe we should just ignore whatever the market does and proceed through political/administrative means to bring about the desired rate of output.

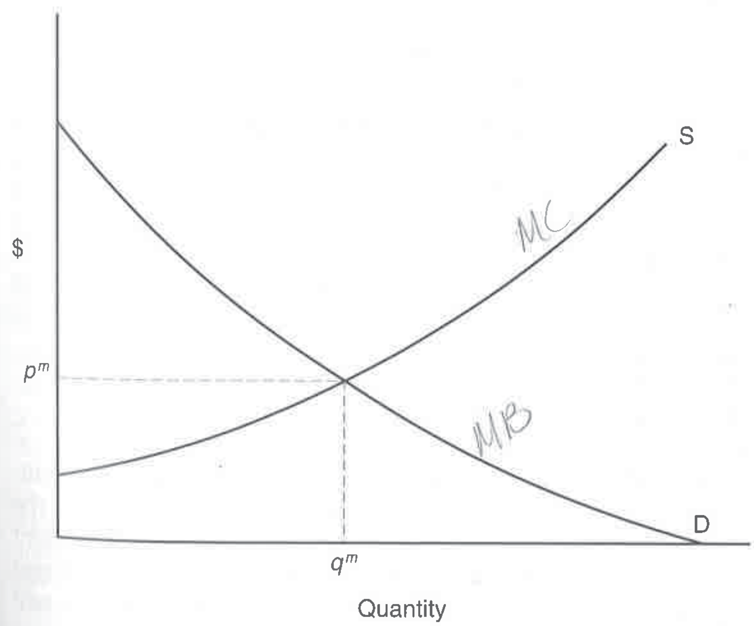
The short answer to this is that as a nation we are in fact committed to a market-based economy. For all its faults, a market system will normally produce better economic results overall than any other system. Those who doubt this need only look at the environmental horror stories uncovered in the countries of Eastern Europe following the Communist era. Of course, it needs to be remembered that although our system is "market based," we do not necessarily have to accept whatever results it yields. The results are acceptable only if they are reasonably efficient and equitable. We will find that in the case of environmental quality, market institutions are not likely to give us results that are socially efficient.

The slightly longer answer to the question is that the market system contains within it certain incentive structures that in many cases can be harnessed toward the objective of improved environmental quality. One of these is the cost-minimizing incentive that stems from the competitive process. Another is the incentive provided through the rewards that may be reaped through initiative in finding better, that is, less expensive, technical and organizational means of production. It will be far more effective in many cases to take advantage of these incentives than to try to do away with them. By altering them so that they take environmental values into account, the market system will yield more-effective results than if we tried to jettison the whole system and adopt a different set of institutions.

A **market** is an institution in which buyers and sellers of consumer goods, factors of production, and so on, carry out mutually agreed-upon exchanges. When they buy or sell in a market, people naturally look for the best terms they can get. Presumably buyers would like to pay a low price whereas sellers would prefer high prices. What brings all these conflicting objectives into balance is the adjustment of prices on the market.

Figure 4.2 shows a simple market model. Buyers' desires are represented by the **demand curve,** labeled D; it shows the quantity of the good that buyers

**FIGURE 4.2** **The Market Model**



would buy at different prices. It has the typical downward slope; the higher the price, the lower the quantity demanded, and vice versa. Underlying the demand curve are such factors as consumer tastes and preferences, the number of potential consumers in the market, and consumer income levels.

The curve labeled S is the **supply curve,** which shows the quantity of the good that suppliers would willingly make available at different prices. It is upward sloping; higher prices represent greater incentives for suppliers, and, therefore, larger quantities are supplied, and vice versa. The main factors affecting the height and shape of the supply curve are production costs. These, in turn, are related to the prices of inputs used in the production of this item and the level of technology inherent in the production process.

It is important to keep in mind that the demand and supply curves represent possibilities, or alternatives. During any particular time, only one quantity of a good can change hands, and sellers and buyers can be on only one point of their supply and demand curves, respectively. It is easy to see that there is only one price at which the quantity demanded by buyers is consistent with the quantity that sellers will make available. That is the price where the two curves intersect, marked $p'''$. Similarly, the total quantity that buyers and sellers will exchange at this price is labeled $q'''$.

For the market to work effectively, there must be competition among sellers and among buyers. None can be large enough that their own performance affects market prices or powerful enough that they can control how the market performs. Price must be allowed to adjust freely so that it can "discover" the quantities that bring buyers and sellers into balance. At prices higher than $p'''$, sellers will attempt to supply more than buyers want. In a surplus situation such as this, competition among sellers forces prices downward. If prices are temporarily lower than $p'''$, a shortage develops and competition among buyers will force the price to adjust upward. At the equilibrium, quantity demanded equals quantity supplied.

It is important to look at it also from the other direction. At the quantity $q'''$ there is an equality between the marginal willingness to pay by consumers for an additional unit of the item and the marginal costs of producing the item. These are equal at the value of $p'''$. If price and quantity are allowed to adjust freely and competition does in fact exist, an equality will arise through the normal interaction of buyers and sellers, between the marginal valuation that consumers have for a good (their marginal willingness to pay) and the cost of making available another unit of the good (the marginal cost of production).

## Markets and Social Efficiency

The next question is whether markets ordinarily produce results that are efficient from the standpoint of society. Compare Figures 4.1 and 4.2. They look the same, but there is actually a big difference. The first shows a socially efficient rate of output for a particular item; the second shows the rate of output and price that would prevail on a competitive market for that item. Are these two

pical downward slope; the higher
and vice versa. Underlying the de-
stes and preferences, the number of
umer income levels.
, which shows the quantity of the
available at different prices. It is
eater incentives for suppliers, and,
d vice versa. The main factors af-
urve are production costs. These, in
in the production of this item and
uction process.
emand and supply curves represent
ticular time, only one quantity of a
ers can be on only one point of their
is easy to see that there is only one
uyers is consistent with the quantity
rice where the two curves intersect,
buyers and sellers will exchange at

must be competition among sellers
ough that their own performance
at they can control how the market
freely so that it can "discover" the
o balance. At prices higher than $p^m$,
buyers want. In a surplus situation
rces prices downward. If prices are
lops and competition among buyers
e equilibrium, quantity demanded

e other direction. At the quantity $q^m$
willingness to pay by consumers for
rginal costs of producing the item.
and quantity are allowed to adjust
an equality will arise through the
between the marginal valuation that
l willingness to pay) and the cost of
d (the marginal cost of production).

## cy

inarily produce results that are effi-
are Figures 4.1 and 4.2. They look the
e. The first shows a socially efficient
econd shows the rate of output and
market for that item. Are these two



rates of output, labeled $q^e$ and $q^m$, likely to be the same in the real world? The answer is yes *if*, and it is a big *if*, the market demand and supply curves, as pictured in Figure 4.2, are the same as the marginal cost and willingness-to-pay curves shown in Figure 4.1. Here is the nub of the problem: When environmental values are concerned, there are likely to be very substantial differences between market values and social values. This is called **market failure**, and it will often call for public intervention, either to override the markets directly or to rearrange things so that they will work more effectively.

In the rest of this chapter we will discuss the performance of markets when matters of environmental quality are involved. There are two phenomena to account for, one on the supply side and the other on the demand side. Environmental effects can drive a wedge between normal market supply curves and true marginal social cost curves. On the other side of the market, environmental effects can create a difference between market demands and true social marginal willingness to pay. On the supply side the problem is "external costs", whereas on the demand side the problem is "external benefits."

## External Costs

When entrepreneurs in a market economy make decisions about what and how much to produce, they normally take into account the price of what they will produce and the cost of items for which they will have to pay: labor, raw materials, machinery, energy, and so on. We call these the **private costs** of the firm; they are the costs that show up in the profit-and-loss statement at the end of the year. Any firm, assuming it has the objective of maximizing its profits, will try to keep its production costs as low as possible. This is a worthwhile outcome for both the firm and society because inputs always have opportunity costs; they could have been used to produce something else. Furthermore, firms will be alert to ways of reducing costs when the relative prices of inputs change. For example, we know that during the U.S. energy "crisis" of the 1970s, when energy inputs became much more expensive, firms found ways of reducing energy inputs by using more energy-efficient machinery, changing operating procedures, and so on.

In many production operations, however, there is another type of cost that, while representing a true cost to society, does not show up in the firm's profit-and-loss statement. These are called **external costs.** They are external because, although they are real costs to some members of society, firms do not normally take them into account when they go about making their decisions about output rates. Another way of saying this is that these are costs that are external to firms but internal to society as a whole.[2]

[2] External costs are sometimes called third-party costs. The first two parties are, respectively, the producer and the consumer. So a third-party cost is one that is inflicted on people who are not directly involved in the economic transactions between buyers and sellers. It is also sometimes called a spillover effect.

One of the major types of external cost is the cost inflicted on people through environmental degradation. An example is the easiest way to see this. Suppose a paper mill is located somewhere on the upstream reaches of a river and that, in the course of its operation, it discharges a large amount of wastewater into the river. The wastewater is full of organic matter that arises from the process of converting wood to paper. This waste material gradually is converted to more benign materials by the natural assimilative capacity of the river water, but, before that happens, a number of people downstream are affected by the lower quality of water in the river. Perhaps the waterborne residuals reduce the number of fish in the river, affecting downstream fishers. The river also may be less attractive to look at, affecting people who would like to swim or sail on it. Worse, the river water perhaps is used downstream as a source of water for a public water supply system, and the degraded water quality means that the town has to engage in more costly treatment processes before the water can be sent through the water mains. All of these downstream costs are real costs associated with producing paper, just as much as the raw materials, labor, energy, and so on, used internally by the plant. But from the mill's standpoint, these downstream costs are **external costs.** They are costs that are borne by someone other than the people who make decisions about operating the paper mill. At the end of the year the profit-and-loss statement of the paper mill will contain no reference to these real downstream external costs.

If rates of output are to be socially efficient, decisions about resource use must take into account both types of costs: the private costs of producing paper plus whatever external costs arise from adverse environmental impacts. In terms of full social cost accounting:

Social costs = Private costs + External (environmental) costs

This is pictured in Figure 4.3. The top panel shows the relationship between the rate of paper production and the occurrence of these downstream external costs. It shows that the marginal external costs increase as paper production increases. The bottom panel shows several things. It shows the demand curve for paper and the marginal private costs of producing paper. The intersection of these occurs at a price of $p^m$ and a quantity of $q^m$. This is the price and quantity that would arise in a competitive market where producers pay no attention to external costs. But marginal social costs are in fact higher, as shown, because they contain both the marginal private costs and the marginal external costs.[3] Thus, the full socially efficient rate of output is $q^*$ and the associated price is $p^*$.

Compare the two rates of output and the two prices. The market output is too high compared to the socially efficient rate of output. In addition, the market price is too low compared to the socially efficient price. It's not hard to understand the reason for this. In considering just its private costs, the firm is essentially using a productive input it is not paying for. What is this unpaid input? The services of the river, which provides the firm with a cheap way to

[3] Note that MEC is zero below a certain quantity. The graph is drawn under the assumption that a **threshold** exists: a quantity of paper production below which there are no external costs.

e cost inflicted on people through
e easiest way to see this. Suppose
tream reaches of a river and that,
large amount of wastewater into
ter that arises from the process of
al gradually is converted to more
apacity of the river water, but, be-
stream are affected by the lower
rborne residuals reduce the num-
fishers. The river also may be less
vould like to swim or sail on it.
stream as a source of water for a
ed water quality means that the
processes before the water can be
vnstream costs are real costs asso-
the raw materials, labor, energy,
from the mill's standpoint, these
e costs that are borne by someone
bout operating the paper mill. At
ent of the paper mill will contain
al costs.
ent, decisions about resource use
e private costs of producing paper
verse environmental impacts. In

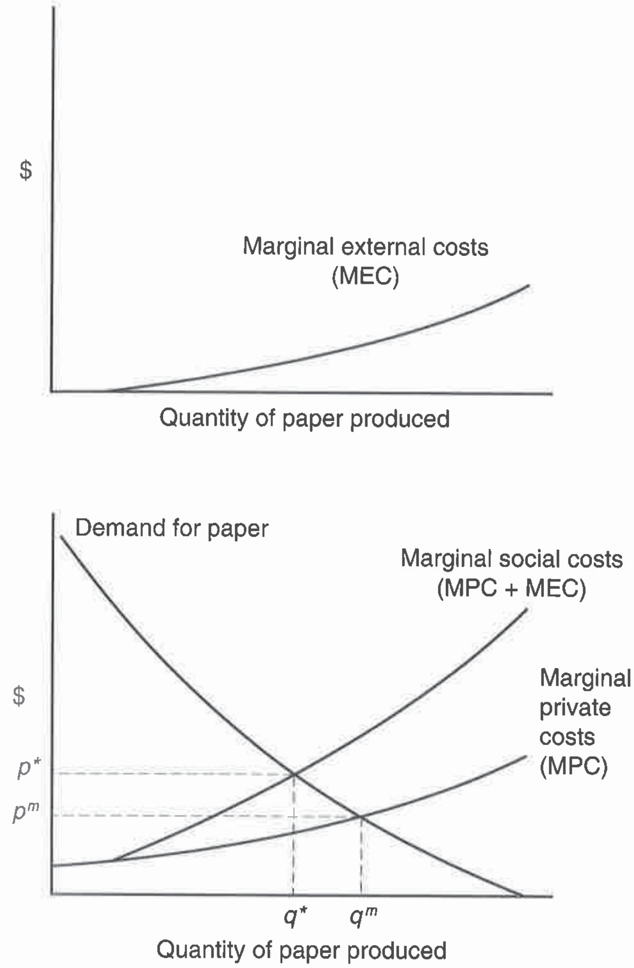
al (environmental) costs

hows the relationship between the
ce of these downstream external
s increase as paper production in-
gs. It shows the demand curve for
ducing paper. The intersection of
f $q^m$. This is the price and quantity
ere producers pay no attention to
in fact higher, as shown, because
and the marginal external costs.[3]
is $q^*$ and the associated price is $p^*$.
two prices. The market output is
rate of output. In addition, the
ally efficient price. It's not hard to
g just its private costs, the firm is
ot paying for. What is this unpaid
ides the firm with a cheap way to

raph is drawn under the assumption that a
which there are no external costs.



**FIGURE 4.3** **External Costs and Market Outcomes**




dispose of its production residuals. Although it may be cheap for the firm to do this, it may not be cheap to society; in fact, in this case we have costs being inflicted on downstream users that are being overlooked by the paper mill. So the private market system in this case produces too much paper at too low a price compared to socially efficient results.

Most of the cases of environmental destruction are related to external costs of one type or another. As a real-world example of external costs, consider the data shown in Example 4.1. It shows some results from a study of the environmental externalities stemming from electricity production. The external effects of a power plant differ according to the technology it uses (nuclear, coal, etc.) and its location relative to centers of population. The numbers in the example show estimates of external costs for a pulverized coal steam plant assumed to be located in the middle of New York City.

## Summary of Estimated External Costs, by Type, for a 300-Megawatt Pulverized Coal Steam Power Plant Assumed to be Located in New York City

**EXAMPLE 4.1**

| Externality Group and Source Group | $/Average Residential Customer[1] |
|---|---|
| Air | |
| Lead | $ 2.27 |
| Mercury | 0.00 |
| Nitrogen oxides | 0.38 |
| Particulates ($PM_{10}$) | 16.22 |
| Radioactivity | 0.00 |
| Sulfur oxides | 2.93 |
| Toxics | 0.03 |
| Air subtotal | 21.83 |
| Water | |
| Chemicals | 0.06 |
| Consumption | 0.00 |
| Toxics in ash | 0.01 |
| Water subtotal | 0.08 |
| Land/waste | |
| Land use/noise/terrestrial | 0.25 |
| Volume/land use | 0.03 |
| Land/waste subtotal | 0.27 |
| Total external costs | $22.18 |

[1] Based on 4,303 kilowatt hours per year for the average residential customer of Consolidated Edison, as reported in "Facts, 1999," at **www.coned.com.** Numbers may not add to totals because of rounding.

**Source:** Based on A. Myrick Freeman III: *The Environmental Cost of Electricity: An Exercise in Pricing the Environment,* Bowdoin College, Economics Department Working Paper 95–116, May 1995.

Note that external costs arising from air pollution account for about 98 percent of the total; that is, costs in terms of water and land pollution are relatively small. In addition, among the various airborne emissions, particulate matter is by far the most serious external cost of generating electricity in this situation.

There are many other types of external costs. Users of chemicals emit toxic fumes that affect people living in the vicinity; developers build on land without taking into account the degradation of the visual environment of local inhabitants; and so on. Nor are businesses the only ones responsible for external environmental costs. When individuals drive their automobiles, exhaust gases add to air pollution, and when they dispose of solid waste materials (e.g., old paint cans), they may affect the quality of the local environment.

Most, but not all, environmental externalities are expressed through physical linkages among parties involved—that is, polluter and people damaged. The simplest is where there are just two parties involved: one polluter and one person suffering damages. An upstream pulp mill and a downstream firm that uses the river water in its production operations are an example. There are cases of single polluters and multiple damaged parties, such as a power plant that emits $SO_2$ affecting a group of community residents living downwind. Other cases involve multiple polluters but only one damaged party, such as the runoff from many farms that affects a local water supply system. Finally, there are many cases where both polluters and parties damaged are many in number. An example of this is urban air pollution stemming from automobile emissions: Each driver is both a producer and a recipient of the externality. The same is true of global phenomena, such as the greenhouse effect.

Some externalities do not involve physical linkages. Degradation of the scenic environment through thoughtless land development is an example. In addition, some externalities involve neither physical linkages nor close proximity. People in one part of a country, for example, may feel loss when those in another region cause damage to an important environmental resource, such as a unique species of animal or plant.

This brings up a problem that we will state but not solve. What is the limit, if any, to be placed on external damage that people may legitimately claim? I suffer damages when someone in my vicinity plays her stereo too loudly, but can I legitimately claim that I suffer damages if, for example, she adopts a lifestyle with which I don't agree? If people in Boston pollute the waters of Boston harbor, may residents in California claim that they have been damaged? If residents of New Jersey thin out the suburban deer population in order to save their flower gardens, may people in Chicago justifiably claim that they have been damaged?

The answer to these questions hinges on the notion of willingness to pay. In this approach, whether someone has or has not been affected by another action hinges on their willingness to pay to have that action changed. If people in New York are willing to pay to preserve clean air in Tokyo, then this is evidence that air quality in Tokyo affects the welfare of people in New York. If people in Chicago are not willing to pay anything to clean up the Ohio River, we conclude that the water quality of that river has no effect on the welfare of people in Chicago. The presence or absence of willingness to pay, in other words, is the economic index of whether an action may be said to affect somebody.

## Open-Access Resources

One source of external costs has been widely studied by environmental economists (as well as natural resource economists): open-access resources. An

**open-access resource** is a resource or facility that is open to uncontrolled access by individuals who wish to use the resource. A classic example is an ocean fishery in which anyone willing to buy a boat and take up fishing is free to do so. Other examples are a pasture that is open to anyone to graze animals, a forest where anyone may go and cut wood, or a public park open to free access.

In these situations we have, in effect, problems in property rights—their definition, distribution, and/or enforcement. If someone owns a pasture or a forest, he or she will presumably keep out encroachers, or perhaps charge them for use of the resource or otherwise control their rate of access. But when a resource or facility is open to unrestricted access, there is no way of ensuring that its rate of use is kept to the level that will maximize its overall value.[4]

To understand this, consider the following example. Suppose there are four similar firms situated on a lake. The firms use the water of the lake in producing their output and discharge emissions back into the lake. Because of the emissions, each firm must treat the water taken from the lake before it uses the water in production. The treatment costs of each firm depend on the ambient quality of the lake, which of course depends on the total emissions of the four firms. Suppose that the cost of intake water treatment is currently $40,000 per year for each firm. A new firm is contemplating starting operations on the lake. If it adds its untreated emissions to those of the current four, it will make ambient water quality worse and drive the cost of water treatment for each firm up to $60,000 per year. When the fifth firm makes its location and production decisions, it will take into account its various operating costs, which will include the $60,000 per year of water treatment costs. But the total social water-related costs of the firm's decisions are higher. There are also external costs inflicted on the other four firms, amounting to $20,000 each of added water treatment costs if the fifth firm locates on the lake. The social marginal costs of water supply when the new firm locates on the lake are $140,000, consisting of $60,000 of internal costs of the new firm plus $80,000 ($20,000 × 4) of external costs inflicted on firms already on the lake. These are often called open-access externalities because they result from the fact that the firms have uncontrolled access to the lake.

We have focused on the externalities flowing from the fifth firm's decisions, but everything is symmetrical in the sense that we could say exactly the same thing about each of the other firms. Each firm will make its decisions without regard to the external costs inflicted on other firms. It is this reciprocal nature of these externalities that distinguishes them from the type we talked about before (e.g., the pulp mill upstream inflicting external costs on people downstream), but the effect is the same: Externalities that lead to rates of output that are too high compared to socially efficient rates.

As another example of an open-access problem, consider a road that is open to access by anyone desiring to use it. A road is not a natural resource but a

[4] This is what is involved in the "tragedy of the commons," as it was popularly termed by Garrett Hardin in "Tragedy of the Commons," *Science*, Vol. 162, December 13, 1968, pp. 1243–1248. His example was an open-access pasture on which all farmers had the right to pasture their sheep.

**TABLE 4.1 Travel Times Related to the Number of Cars on the Road**

| Number of Cars | Average Travel Time between A and B |
|---|---|
| 10 | 10 |
| 20 | 10 |
| 30 | 10 |
| 40 | 11 |
| 50 | 12 |
| 60 | 14 |
| 70 | 18 |
| 80 | 24 |

person-made facility. But the essence of the uncontrolled access problem is identical, and perhaps it is easier to understand with this particular example. It uses very simplifying assumptions in order to highlight the basic issues. There is a road connecting two points—Point A and Point B. The figures in Table 4.1 show the average travel time it takes to get from Point A to Point B along this road, as a function of the number of motorists using the road. Thus, for example, if there are just 10 travelers on the road, it takes 10 minutes to get from A to B (we assume a speed limit that is enforced). Likewise, when there are either 20 or 30 motorists on the road, the average travel time is still 10 minutes, but when the traffic increases to 40 travelers, the average travel time increases to 11 minutes. This is because of congestion; cars begin to get in each other's way and average speeds drop. As the number of motorists continues to increase, the congestion increases, thus driving up travel times even more.

Now suppose you are considering using this road to go from A to B and that there are already 50 cars using it. Suppose, furthermore, that you have an alternative route that will take you 18 minutes. Assume that you know the state of the traffic and the resulting travel times. Because taking the given road will save you 4 minutes over the alternative, your individual decision would be to use the road. But from the standpoint of "society," in this case consisting of you plus all the other motorists on the road, this is not efficient. When you enter the highway on which there are already 50 cars, the added congestion causes an increase in average travel times of 2 minutes to the people already using the road. Thus, your 4-minute individual savings is offset by added travel costs of 100 minutes (50 cars times 2 minutes per car) on the part of the other motorists, meaning that if all minutes are treated as equally valuable, there is a net social loss of 96 minutes when you decide to use the road.

The problem arises because there is uncontrolled access to the road, and in using it people may inflict external costs on others in the form of added congestion and higher travel times. The same kind of effect holds when a fisher enters a fishery; in catching a portion of the stock, he leaves fewer to be caught by other fishers. When one farmer puts animals on a common pasture, he or she reduces the forage available to other herds on that pasture. When one person cuts wood from a communal forest, she leaves fewer trees for other users and makes it more difficult for them to supply themselves with wood. We can see that this is related to the notion of external costs. The added costs that one user of a common-property resource inflicts on other users of that resource are in fact costs that are external to that user but internal to the whole group of users. When a single individual is making a decision about whether and how much to utilize a common-property resource, she takes into account the costs and benefits that impinge directly on her. Some people might also altruistically take into account the common-property externalities they inflict on others, but most will not. The result will be, as it was with the road example, a rate of use that is higher than what is called for on grounds of social efficiency.

Thus, when external costs are present, private markets will not normally produce quantities of output that are socially efficient. This market failure may justify public policy to help move the economy toward efficiency. This may be done sometimes by changing rules, such as property rights rules, so that the market will function efficiently. Other cases may call for more direct public intervention. We will take up these matters again in Section 4. We must now move to the demand side of the market and consider another important source of market failure, that of external benefits.

## External Benefits

An **external benefit** is a benefit that accrues to somebody who is outside, or external to, the decision about consuming or using the good or resource that causes the externality. When the use of an item leads to an external benefit, the market willingness to pay for that item will understate the social willingness to pay. Suppose a quieter lawn mower would provide $50 a year of extra benefits to me if I were to buy it. This is therefore the maximum that I would be willing to pay for this machine. But suppose my use of the new lawn mower would create $20 of added benefits to my neighbor because of reduced noise levels in the course of the year. These $20 of benefits to the neighbor are external benefits for me. I make my purchasing decision on the basis of benefits accruing only to me. Thus, my marginal willingness to pay for a quieter lawn mower is $50, whereas the social marginal benefits (where "society" in this case includes just me and my neighbor) is $70 (my $50 and her $20).

As another example of an external benefit, consider a farmer whose land is on the outskirts of an urban area. The farmer cultivates the land and sells his produce to people in the city. Of course, the farmer's main concern is the income he can derive from the operation, and he makes decisions about inputs and outputs according to their effect on that income. But the land kept in agriculture produces several other benefits, including a habitat for birds and other small animals and scenic values for passers-by. These benefits, although internal from the standpoint of society, are external from the standpoint of the farmer. They don't appear anywhere in his profit-and-loss position; they are external benefits of his farming decisions. In this case the agricultural value of the land to the farmer understates the social willingness to pay to have the land in agriculture.

Many goods do not involve external benefits. Indeed, when economists discuss the rudiments of supply and demand, the examples used are normally simple goods that do not create this complication. Farmers produce and supply so many thousand cantaloupes; individual and market demand curves for cantaloupes are easy to comprehend. If we want to know the total number of cantaloupes bought, we can simply add up the number bought by each person in the market. Each person's consumption affects no one else. In this case the market demand curve will represent accurately the aggregate marginal willingness to pay of consumers for cantaloupes. But in cases involving external benefits, this no longer holds. We can perhaps best see this by considering a type of good that inherently involves large-scale external benefits, what economists have come to call "public goods."

## Public Goods

Consider a lighthouse. This is a service provided to mariners at sea so that they can locate themselves and avoid running aground at night. But the lighthouse has an interesting technical characteristic: If its services are made available to one mariner at sea, they immediately become available to all others in the vicinity. Once the services are made available to one person, others cannot be excluded from making use of the same services. This is the distinguishing characteristic of a **public good.** It is a good that, if made available to one person, automatically becomes available to others.

Another example of a public good is a radio signal. Once a radio station broadcasts a signal, it is available to anybody who has a receiver. Each individual can listen to the broadcast without diminishing its availability to all other people within range of the station. Note carefully that it is not the ownership of the supplying organization that makes a public good public. Lighthouses are usually publicly owned, but radio stations, at least in the United States, are typically privately owned. A public good is distinguished by the technical nature of the good, not by the type of organization making it available.

We are interested in public goods because environmental quality is essentially a public good. If the air is cleaned up for one person in an urban area, it is automatically cleaned up for everybody else in that community. The benefits, in other words, accrue to everyone in the community. Private markets are likely to undersupply public goods, relative to efficient levels. To see why, let's take another very simple example: a small freshwater lake, the shores of which have three occupied homes. The people living in the houses use the lake for recreational purposes, but, unfortunately, the water quality of the lake has been contaminated by an old industrial plant that has since closed. The contaminant is measured in parts per million (ppm). At present the lake contains 5 ppm of this contaminant. It is possible to clean the water by using a fairly expensive treatment process. Each of the surrounding homeowners is willing to pay a certain amount to have the water quality improved. Table 4.2 shows these individual marginal willingnesses to pay for integer values of water quality. It also shows the aggregate marginal willingness to pay, which is the sum of the individual values.

**TABLE 4.2 Individual and Aggregate Demand for Lowering Lake Pollution**

| Level of Contaminant (ppm) | Marginal Willingness to Pay ($ per year) Homeowner A | Homeowner B | Homeowner C | Aggregate MWP | Marginal Cost of Cleanup |
|---|---|---|---|---|---|
| 4 | 110 | 60 | 30 | 200 | 50 |
| 3 | 85 | 35 | 20 | 140 | 65 |
| 2 | 70 | 10 | 15 | 95 | 95 |
| 1 | 55 | 0 | 10 | 65 | 150 |
| 0 | 45 | 0 | 5 | 50 | 240 |

The table also shows the marginal cost of cleaning up the lake, again just for integer values of water quality. Note that marginal cost is increasing; as the lake becomes cleaner, the marginal cost of continued improvement increases. Marginal cost and aggregate marginal willingness to pay are equal at a water quality of 2 ppm. At levels less than this (higher ppm), aggregate marginal willingness to pay for a cleaner lake exceeds the marginal cost of achieving it. Hence, from the standpoint of these three homeowners together, improved water quality is desirable, but at quality levels better than 2 ppm total willingness to pay falls below marginal costs. Thus, 2 ppm is the socially efficient level of water quality in the lake.

This is depicted graphically in Figure 4.4. The top three panels show the marginal willingness to pay by each of the three homeowners. When summing individual demand curves for private goods, we could add together the individual quantities demanded at each price to get the aggregate quantity demanded. But with a public good people are, in effect, consuming the same units, so we must add together the individual marginal willingness to pay at each quantity to get the aggregate demand function, as shown in Figure 4.4. At a water-quality level of 3 ppm, for example, the marginal willingnesses to pay are, respectively, $85, $35, and $20 for individuals A, B, and C. Thus, the aggregate marginal willingness to pay at this level of water quality is $140. The bottom panel of the graph shows the aggregate marginal willingness-to-pay/demand function labeled D, the marginal cost function (MC), and the efficient level of water quality.

Having identified the efficient level of water quality, could we rely on a competitive market system, where entrepreneurs are on the alert for new profit opportunities, to get the contaminant in the lake reduced to that level? Suppose a private firm attempts to sell its services to the three homeowners. The firm goes to person A and tries to collect an amount equal to that person's true willingness to pay. But that person will presumably realize that once the lake is cleaned up, it is cleaned up for everybody no matter how much each homeowner actually contributed. So A may have the incentive to underpay, relative to his true willingness to pay, in the hopes that the other homeowners will contribute enough to cover the costs of the cleanup. Of course, the others may react in the same way. When a public good is involved, each person may have an r Lowering Lake Pollution

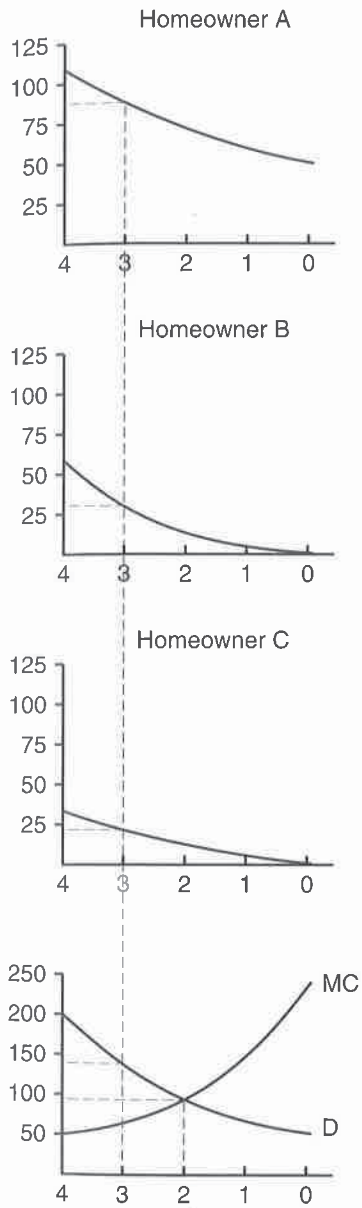
| ar) | | |
|---|---|---|
| vner | Aggregate MWP | Marginal Cost of Cleanup |
| | 200 | 50 |
| | 140 | 65 |
| | 95 | 95 |
| | 65 | 150 |
| | 50 | 240 |

aning up the lake, again just for
nal cost is increasing; as the lake
nued improvement increases.
ess to pay are equal at a water
ppm), aggregate marginal will-
marginal cost of achieving it.
meowners together, improved
better than 2 ppm total willing-
pm is the socially efficient level

top three panels show the mar-
homeowners. When summing
could add together the individ-
aggregate quantity demanded.
nsuming the same units, so we
ngness to pay at each quantity to
n Figure 4.4. At a water-quality
gnesses to pay are, respectively,
us, the aggregate marginal will-
0. The bottom panel of the graph
ay/demand function labeled D,
t level of water quality.
quality, could we rely on a com-
e on the alert for new profit op-
reduced to that level? Suppose
e three homeowners. The firm
equal to that person's true will-
ly realize that once the lake is
matter how much each home-
incentive to underpay, relative
the other homeowners will con-
Of course, the others may react
ved, each person may have an



**FIGURE 4.4** **Aggregate Willingness to Pay for a Public Good**



incentive to **free ride** on the efforts of others. A free rider is a person who pays less for a good than her or his true marginal willingness to pay, that is, a person who underpays relative to the benefits he receives.

Free riding is a ubiquitous phenomenon in the world of public goods, or in fact any good the consumption of which produces external benefits. Because of the free-riding impulse, private, profit-motivated firms will have difficulty covering their costs if they go into the business of supplying public goods.[5] Because of these reduced revenues, private firms will normally *undersupply* goods and services of this type. Environmental quality improvements are essentially public goods. Because we cannot rely on the market system to provide efficient quantities of goods of this type, we must fall back on some type of nonmarket institution involving collective action of one type or another. In the lake example, the homeowners may be able to act together **privately**, perhaps through a homeowners' association, to secure contributions for cleaning up the lake. Of course, the free-rider problem will still exist even for the homeowners' association, but if there are not too many of them, personal acquaintance and the operation of moral pressure may be strong enough to overcome the problem. When there are many more people involved (thousands, or perhaps millions, as there are in many large urban areas), the free-rider problem can be addressed effectively only with more direct governmental action. This opens up the huge topic of **public policy** for environmental quality, a topic we will spend much more time discussing throughout the rest of the book.

## Summary

The main goal in this chapter was to discuss the operation of private markets and then apply the market model to situations in which environmental quality is an issue. Markets are places where buyers and sellers interact over the quantities and prices of particular goods or services. Buyers' desires are represented by the aggregate demand curve, which shows the quantities demanded at alternative prices. Sellers' supply capabilities are represented by supply curves, which ultimately are based on underlying production costs and show quantities that would be made available at alternative prices. The intersection of supply and demand curves shows the unique quantity and price that can simultaneously satisfy both buyers and sellers. For many types of goods and services, market outcomes (output and price levels) also may be the outcomes that are socially efficient. Outcomes that are socially efficient are those in which aggregate marginal willingness to pay in society is equal to aggregate marginal social costs of production. When market results are not socially efficient, we speak of market failures.

[5] This sentence emphasizes the point made earlier: It is the technical nature of the good that makes it a public or private good, not whether the organization providing it is public or private. A lighthouse (a public good) might be built and operated by a private firm; insurance (a private good) might be provided by a public agency.

# Chapter 5

# The Economics of Environmental Quality

In the preceding chapter we concluded that the market system, left to itself, is likely to malfunction when matters of environmental pollution are involved. That is to say, it will not normally produce results that are socially efficient. This brings us to the **policy question:** If we do not like the way things are currently turning out, what steps should be undertaken to change the situation?[1]

The policy problem includes a number of closely related issues. One of the first is that of identifying the most appropriate level of environmental quality we ought to try to achieve. Another is how to divide up the task of meeting environmental quality goals. If we have many polluters, how should we seek to allocate among them an overall reduction in emissions? Another issue is the question of how the benefits and costs of environmental programs are distributed across society and whether this distribution is appropriate. In this chapter we take up these issues on a conceptual basis; in subsequent chapters we will look at specific policy alternatives.

Before developing a simple policy model, we need to stress again that effective public policy depends on good information on how economic and environmental systems actually work. This might be called the scientific basis of environmental policy—that is, the study of how firms and consumers normally make decisions in the market economy, how residuals are emitted into the natural environment, and the ways in which these residuals behave in that environment to produce human and nonhuman damages. Thousands of scientists have worked and continue to work on these issues to clarify these diverse linkages. Great effort will continue to be needed to expand the scientific base on which to develop environmental policy.

[1] This goes back to the distinction made earlier between positive and normative economics (see p. 3). Explaining why there is a certain amount of $SO_2$ in the air at any particular time is a question of positive economics; deciding what best to do about it is a case of normative economics.

of
Quality

he market system, left to itself, is onmental pollution are involved. ults that are socially efficient. This like the way things are currently to change the situation?[1]
closely related issues. One of the te level of environmental quality divide up the task of meeting en- polluters, how should we seek to emissions? Another issue is the ironmental programs are distrib- ion is appropriate. In this chapter s; in subsequent chapters we will

e need to stress again that effective on how economic and environ- called the scientific basis of envi- firms and consumers normally residuals are emitted into the nat- ese residuals behave in that envi- damages. Thousands of scientists ssues to clarify these diverse link- to expand the scientific base on

positive and normative economics
$O_2$ in the air at any particular time is
do about it is a case of normative



## Pollution Control—A General Model

Diverse types of environmental pollutants obviously call for diverse types of public policy, but in order to build up the required policy analyses it is better to start with one very simple model that lays out the fundamentals of the policy situation. The essence of the model consists of a simple **trade-off** situation that characterizes all pollution-control activities. On the one hand, reducing emissions reduces the damages that people suffer from environmental pollution; on the other hand, reducing emissions takes resources that could have been used in some other way.

To depict this trade-off consider a simple situation where a firm (e.g., a pulp mill) is emitting production residuals into a river. As these residuals are carried downstream, they tend to be transformed into less damaging chemical constituents, but before that process is completed the river passes by a large metropolitan area. The people of the area use the waters of the river for various purposes, including recreation (boating, fishing) and as a source for the municipal water supply system. When the river becomes polluted with industrial waste, the people downstream are damaged by the disruption of these and other services provided by the river. One side of the trade-off, then, is the **damages** that people experience when the environment is degraded.

Upstream, the offending pulp mill could reduce the amount of effluent put in the river by treating its wastes before discharge, as well as by recycling certain materials that currently just run out of the discharge pipe. This act of reducing, or abating, some portion of its wastes will require resources of some amount, the costs of which will affect the price of the paper it produces.[2] These **abatement costs** are the other side of the basic pollution-control trade-off.

## Pollution Damages

By **damages** we mean all the negative impacts that users of the environment experience as a result of the degradation of that environment. These negative impacts are of many types and, of course, will vary from one environmental asset to another. In the river pollution example, damages were to recreators, who could no longer use the river or who suffered a higher chance of picking up waterborne diseases, and to all the city dwellers who had to pay more to treat the water before they could put it into the public water mains.

Air pollution produces damage through its impacts on human health. Excess deaths from diseases such as lung cancer, chronic bronchitis, and emphysema are related to elevated levels of various pollutants, such as particulate matter, asbestos fibers, and radon emissions. Air pollution can cause damages through the degradation of materials (all of the important outdoor sculpture from

[2] The word *resources* has a double meaning in economics. On the one hand it is a shorthand way of referring to natural resources. On the other hand, it is more generally used to refer to the inputs that are utilized to produce outputs.

**TABLE 5.1 Estimated Benefits (Reduced Damages) in 2010 from Clean Air Act Reductions of Criteria Pollutants**

| | $ Millions (1990 dollars)* |
|---|---|
| Mortality† | 100,000 |
| Chronic illness | |
| Chronic bronchitis | 5,600 |
| Chronic asthma | 180 |
| Hospitalization | |
| All respiratory | 130 |
| Total cardiovascular | 390 |
| Asthma-related ER visits | 1 |
| Minor illness | |
| Acute bronchitis | 2.1 |
| Upper respiratory symptoms | 19 |
| Lower respiratory symptoms | 6.2 |
| Respiratory illness | 6.3 |
| Moderate/worse asthma | 13 |
| Asthma attacks | 55 |
| Chest tightness, shortness of breath | 11 |
| Work-loss days | 340 |
| MRAD/any of 19‡ | 1,200 |
| Welfare | |
| Decreased worker productivity | 710 |
| Visibility-recreational | 2,900 |
| Agriculture | 550 |
| Acidification | 50 |
| Commercial timber | 600 |
| Aggregate | 110,000 |

*This means that the estimates for 2010 were done in terms of 1990 dollars, that is, they were corrected for anticipated inflation between 1990 and 2010.

†This is the estimated value associated with the reduction in premature mortality.

‡Minor restricted activity days stemming from any of 19 different respiratory symptoms.

**Source:** U.S. EPA, "The Benefits and Costs of the Clean Air Act of 1990 to 2010," *EPA Report to Congress, EPA-410-R-99-001*, Washington, DC, November 1999, p. 102.

Renaissance Florence has had to be put inside to protect it from air pollution) and the deterioration of the visual environment. Table 5.1 shows the range of impacts produced by the major air pollutants in the United States. It is in terms of the **damages reduced** (i.e., benefits) by the Clean Air Act.

Besides damage to human beings, environmental destruction can have important impacts on various elements of the nonhuman ecosystem. Some of these, such as the destruction of genetic information in plant and animal species driven to extinction, will ultimately have important implications for humans. Estimating environmental damages is one of the primary tasks facing environmental scientists and economists, and we will devote Chapter 7 to a discussion of this problem.

es) in 2010 from Clean Air Act

| $ Millions (1990 dollars)* |
|---|
| 100,000 |
| 5,600 |
| 180 |
| 130 |
| 390 |
| 1 |
| 2.1 |
| 19 |
| 6.2 |
| 6.3 |
| 13 |
| 55 |
| 11 |
| 340 |
| 1,200 |
| 710 |
| 2,900 |
| 550 |
| 50 |
| 600 |
| 110,000 |

ollars, that is, they were corrected for anticipated

re mortality.
piratory symptoms.
90 to 2010," *EPA Report to Congress,*

de to protect it from air pollution)
nent. Table 5.1 shows the range of
s in the United States. It is in terms
e Clean Air Act.
mental destruction can have impor-
nhuman ecosystem. Some of these,
n in plant and animal species driven
implications for humans. Estimating
y tasks facing environmental scien-
pter 7 to a discussion of this problem.



## Damage Functions

In general, the greater the pollution, the greater the damages it produces. To describe the relationship between pollution and damage, we will use the idea of a **damage function.** A damage function shows the relationship between the quantity of a residual and the damage that residual causes. There are two types of damage functions.

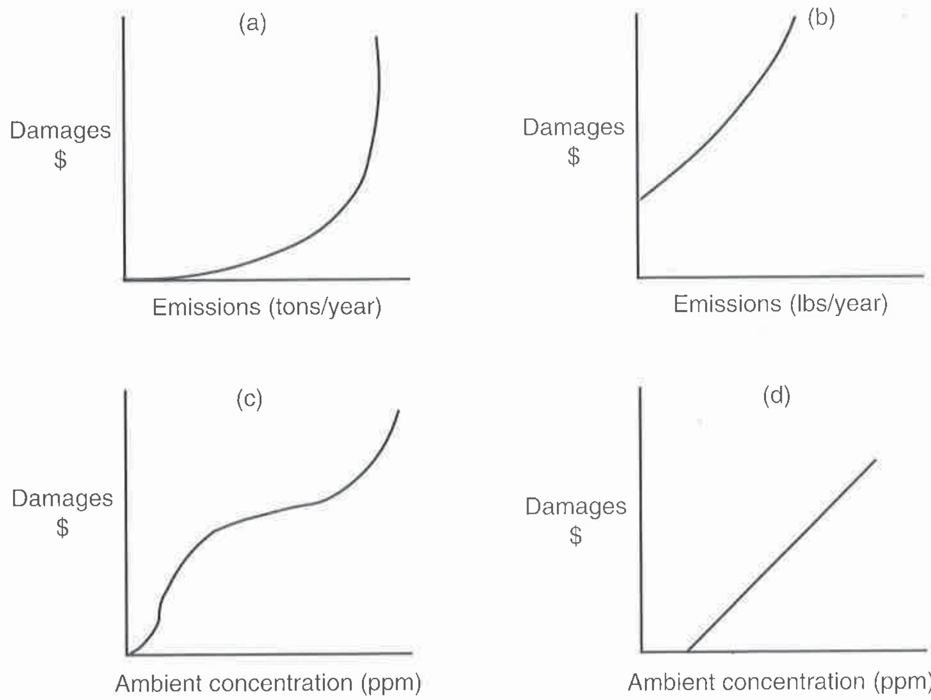
- **Emission damage functions:** These show the connection between the quantity of a residual emitted from a source or group of sources and the resulting damage.
- **Ambient damage functions:** These show the relationship between concentration of particular pollutants in the ambient environment and the resulting damages.

Damage functions can be expressed in a variety of ways, but our primary model will make use of **marginal damage functions.** A marginal damage function shows the **change** in damages stemming from a unit change in emissions or ambient concentration. When necessary, we also can use these relationships to discuss total damages because we know that, graphically, the areas under marginal damage functions correspond to total damages.

The height and shape of a damage function depends on the pollutant and circumstances involved. Several marginal damage functions are depicted in Figure 5.1. The top two are marginal emission damage functions; the horizontal axes measure the quantity of an effluent emitted into the environment during some specified period of time. The exact units (pounds, tons, etc.) in any particular case depend on the specific pollutant involved. The vertical axes measure environmental damages. In physical terms, environmental damage can include many types of impacts: miles of coastline polluted, numbers of people contracting lung disease, numbers of animals wiped out, quantities of water contaminated, and so on. Every case of environmental pollution normally involves multiple types of impacts, the nature of which will depend on the pollutant involved and the time and place it is emitted. To consider these impacts comprehensively we need to be able to aggregate them into a single dimension. For this purpose we use a monetary scale. It is sometimes easy to express damage in monetary units—for example, the "defensive" expenditures that people make to protect themselves against pollution (e.g., heavier insulation to protect against noise). Usually, however, it is very difficult, as we will see.

The marginal emission damage function in panel (a) of Figure 5.1 shows marginal damages increasing only modestly at the beginning but more rapidly as emissions get larger and larger. Work by environmental scientists and economists seems to suggest that this is a typical shape for many types of pollutants, although probably not for all of them. At low levels of emissions, marginal damages may be comparatively small; ambient concentrations are so modest that only the most sensitive people in the population are affected. But when emission levels go higher, damages mount, and at still higher levels of emissions, marginal damages become very elevated as environmental impacts become widespread and intense.

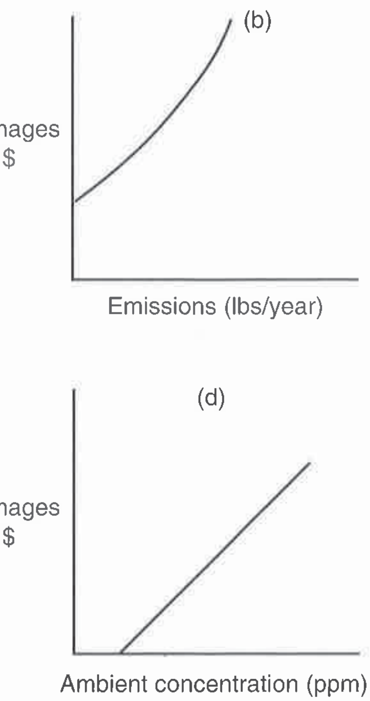
**FIGURE 5.1** Representative Marginal Damage Functions




Panel (b) shows a marginal (emission) damage function that has the same general shape as panel (a) (i.e., it shows increasing marginal damage), but it begins much higher on the vertical axis and rises more sharply. It might represent a toxic substance that has a deadly effect even at very low levels of emission.

The two bottom relationships in Figure 5.1 are marginal ambient damage functions. Whereas the vertical axes have a monetary index of damages, the horizontal axes have an index of ambient concentration, such as parts per million (ppm). Panel (c) shows a complicated function that increases at low concentrations, then tends to level off until much higher concentrations are reached, after which damages increase rapidly. This might apply, for example, to an air pollutant that causes marked damages among particularly sensitive members of society at relatively low concentrations, and among all people at very high concentrations, while in the middle ranges marginal damages do not increase rapidly. Panel (d) demonstrates an ambient marginal damage function that begins to the right of the origin and then increases linearly with ambient concentration.

Panels (a) and (d) illustrate a characteristic that is in fact quite controversial. They have **thresholds;** that is, they have values of emissions or ambient concentrations below which marginal damages are zero. Thus, the pollutant can increase to these threshold levels without causing any increase in damages. As will be seen in chapters to come, the assumed existence or nonexistence of a threshold in the damage functions for particular pollutants has had important impacts on real-world environmental control policies. There have been long, unctions




e function that has the same gen- marginal damage), but it begins harply. It might represent a toxic w levels of emission.

marginal ambient damage func- index of damages, the horizontal uch as parts per million (ppm). eases at low concentrations, then ns are reached, after which dam- ole, to an air pollutant that causes members of society at relatively high concentrations, while in the e rapidly. Panel (d) demonstrates ns to the right of the origin and on.

hat is in fact quite controversial. es of emissions or ambient con- e zero. Thus, the pollutant can ng any increase in damages. As existence or nonexistence of a r pollutants has had important policies. There have been long,



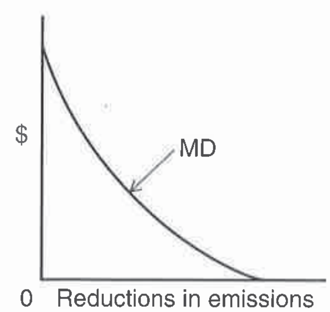
vigorous arguments about whether the damage functions of certain types of pollutants do or do not have thresholds.

## Damage Functions: A Closer Look

We need to look more deeply into the concept of the damage function because it will be used later to express and analyze a variety of different types of pollution problems and public policy approaches. Accordingly, Figure 5.2 shows two marginal emissions damage functions.[3] It is important to remember that, like the demand and supply curves discussed earlier, these are time specific; they show the emissions and the marginal damages for a particular period of time. There are a couple of ways of thinking about this. One is to assume, for purposes of simplicity, that the graph refers to a strictly noncumulative pollutant. Thus, all damages occur in the same period as emissions. A somewhat more complicated assumption is that for a pollutant that cumulates over time, the damage function shows the total value that people place on current and future damages. In Chapter 6 we will discuss this concept more fully.

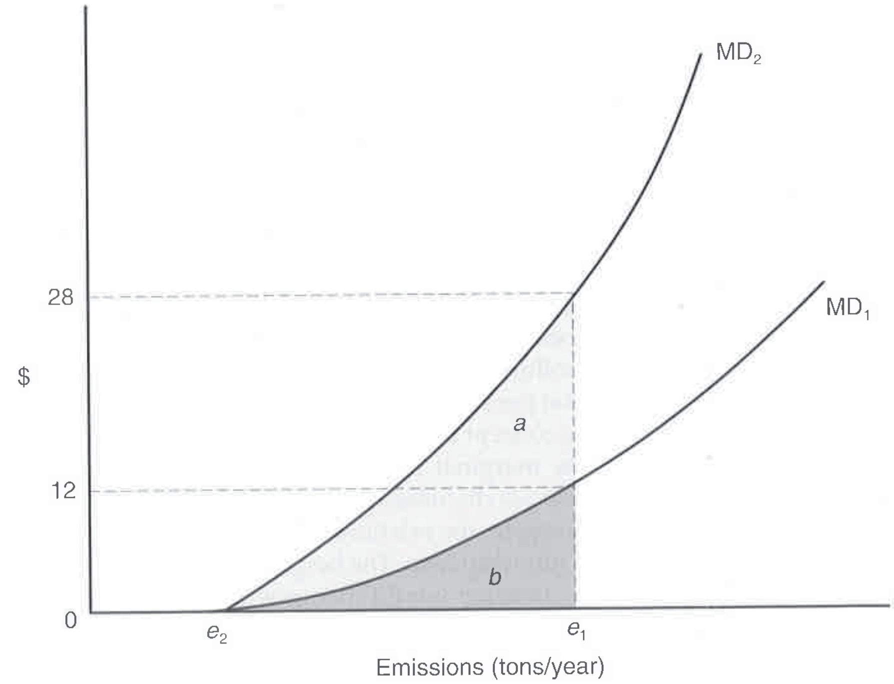
Consider first just one of the marginal damage functions, the lower one labeled $MD_1$. In previous chapters we discussed the relationship between marginal and total quantities, for example, the relationship between marginal and total costs. We have the same relationship here. The height of the marginal damage curve shows how much total damages would change with a small change in the quantity of emissions. When the effluent level is at the point marked $e_1$, for example, marginal damages are $12. That is to say, if emissions were to increase by one ton from point $e_1$, the damages experienced by people exposed to those emissions would increase by $12. By the same token, if emissions decreased by a small amount at point $e_1$, total damages would be reduced by $12. Because the height of

[3] The marginal damage function goes up to the right because the quantity on the *x*-axis is emissions, which start at zero and increase to the right. Reducing pollution is thus going to be a move to the left, and the benefits this produces are shown by the reduction in marginal damages in a leftward move. In some models, however, what is indexed on the horizontal axis is *reductions* from current emission levels. Then a move to the right corresponds to a reduction in pollution, and the marginal damage function appears as a standard marginal benefit function, pictured here.




Of course either approach allows the same analysis. We chose the former (pollution upward to the right) because it means the origin corresponds to no pollution. In our models the rising marginal damage function might suggest to you a rising marginal cost curve. In essence it is, though in this case the marginal cost refers to the marginal cost to society of increasing pollution.

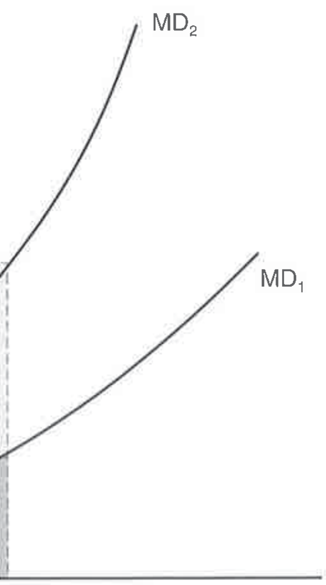
**FIGURE 5.2 Anatomy of a Marginal Damage Function**




the curve, as measured on the $y$-axis, shows marginal damages, the area under the curve between the point where it is zero and some other point, like the one labeled $e_1$, shows the total damages associated with that level of emissions. In the case of marginal damage function $MD_1$ and point $e_1$, total damages are equal to the monetary amount expressed by the triangular area bounded by the $x$-axis, the curve $MD_1$, and the effluent quantity $e_1$. That is area $b$ in Figure 5.2.

What factors might account for the difference between $MD_1$ and $MD_2$ in Figure 5.2? Let us assume that they apply to the same pollutant. For any given level of emissions, marginal damages are higher for $MD_2$ than for $MD_1$. At emission level $e_1$, for example, a small increase in effluent would increase damages by \$12 if the marginal damage function were $MD_1$, but it would increase damages by \$28 if it were $MD_2$. Remember that any damage function shows the impacts of emitting a particular effluent in a particular time and place, so one possible explanation might be that $MD_2$ refers to a situation in which many people are affected by a pollutant, such as a large urban area, whereas $MD_1$ refers to a more sparsely populated rural area—fewer people, smaller damage. One major factor that moves damage functions upward, in other words, is an increase in the number of people exposed to a particular pollutant.

Another possibility that might offer an explanation of why one marginal damage function lies above another is that although they apply to the same group of people, they refer to different time periods. Damage results from ambient pollution, whereas what we have on the horizontal axis is quantity of unction




arginal damages, the area under
some other point, like the one la-
ith that level of emissions. In the
int $e_1$, total damages are equal to
ar area bounded by the $x$-axis, the
area $b$ in Figure 5.2.
e between MD$_1$ and MD$_2$ in Fig-
me pollutant. For any given level
MD$_2$ than for MD$_1$. At emission
would increase damages by $12
it it would increase damages by
e function shows the impacts of
time and place, so one possible
ituation in which many people
an area, whereas MD$_1$ refers to a
ple, smaller damage. One major
in other words, is an increase in
ollutant.
planation of why one marginal
lthough they apply to the same
riods. Damage results from am-
e horizontal axis is quantity of



emissions. The functioning of the environment is what connects these two factors. Suppose the pollutant in question is some sort of material emitted into the air by industrial firms located near an urban area and that the damage functions refer to impacts felt by people living in that area. Marginal damage function MD$_2$ might occur when there is a temperature inversion that traps the pollutant over the city and produces relatively high ambient concentrations. MD$_1$ would be the damage function, however, when normal wind patterns prevail so that most of the effluent is blown downwind and out of the area. Thus, the same emission levels at two different times could yield substantially different damage levels due to the workings of the natural environment.

Because MD$_2$ is above MD$_1$, it corresponds not only to higher marginal damages but also to higher total damages. At emission level $e_1$, total damages are equal to area $b$ when the damage function is MD$_1$, but to area $(a + b)$ when the damage function is MD$_2$.

Having considered the concept of damages, it is now necessary to look at the other side of the trade-off relationship mentioned previously. It is tempting not to do this, to conclude instead that the damage functions themselves give us all the information needed to make decisions about pollution control. One might be tempted to say, for example, that society ought to strive for emission levels around point $e_2$ where marginal damages are zero, or perhaps even the origin, corresponding to a point at which emissions are zero. There may be certain pollutants and situations where the efficient level of emissions is indeed zero. But to determine this we have to look at the other side of the problem: abatement costs. We consider abatement costs after the next section.

## Damages and Uncertainty

The damage functions drawn above give the appearance of being very clear and unambiguous. In the real world, however, they hardly ever are. Usually there is a lot of uncertainty about the connections between pollution emissions and various types of damage: health impacts on humans, ecosystem damages, and so on. When we say "uncertain," we are not implying that pollution actually causes less damage than we might have thought, but rather that the exact amount of damage caused by different levels of pollution is difficult to measure with certainty. Uncertainty in nature affects the relationship between emissions and ambient environmental conditions, and uncertainty in human reaction affects the damages that result. This is exacerbated by the fact that much of the damage can be expected to occur well off in the future, making it difficult to predict with accuracy.

Another factor of importance is the implicit assumption we are making that damage functions are **reversible.** If emissions increase, damages increase; and if emissions decrease, damages will go back to their previous level. This may fit many pollutants: more ozone, more asthma; less ozone, and cases of asthma go back down. But for many pollutants this may not be true. The buildup of global greenhouse gases could perhaps initiate global changes that are essentially irreversible. Even some local changes may be of this type; higher levels of pollution lead to ecosystem changes that take us to new situations from which there is no easy return. For example, once a groundwater aquifer is contaminated, it may never be the same again.

Environmental economists, in cooperation with environmental scientists, epidemiologists, and the like, have worked to develop means by which damage functions can be measured with greater accuracy. In the next few chapters we will look at some of these methods. In this chapter, we will simply use the concept of the damage function to study the essential choices that face society in pollution control decisions.

## Abatement Costs

**Abatement costs** are the costs of reducing the quantity of residuals being emitted into the environment, or of lowering ambient concentrations. Think of the pulp mill located upstream. In its normal course of operation it produces a large quantity of organic wastes. On the assumption that it has free access to the river, the cheapest way to get rid of these wastes is simply to pump them into the river, but the firm normally has technological and managerial means to reduce these emissions. The costs of engaging in these activities are called "abatement costs" because they are the costs of abating, or reducing, the quantity of residuals put into the river. By spending resources on this activity, the pulp mill can abate its emissions; in general, the greater the abatement, the greater the cost.

Abatement costs normally will differ from one source to another, depending on a variety of factors. The costs of reducing emissions of $SO_2$ from electric power plants obviously will be different from the costs of reducing, say, toxic fumes from chemical plants. Even for sources producing the same type of effluent the costs of abatement are likely to be different because of differences in the technological features of the operation. One source may be relatively new, using modern production technology, whereas another may be an old one using more highly polluting technology. In the discussion that follows keep in mind that *abatement* is used with the widest possible connotation and includes all the many ways there are of reducing emissions: changes in production technology, input switching, residuals recycling, treatment, abandonment of a site, and so forth.

### Abatement Cost Functions

We represent this idea graphically using the concept of the **marginal abatement cost** function. The units on the axes are the same as before: quantities of pollutants on the horizontal axis and monetary value on the vertical axis. Marginal emission abatement costs show the added costs of achieving a one-unit decrease in emission level, or alternatively the costs saved if emissions are increased by a unit. On the horizontal axis, marginal abatement cost curves originate at the uncontrolled emission levels, that is, emission levels prior to undertaking any abatement activities. From this origin point, marginal abatement costs show the marginal costs of producing reductions in emissions. Thus, these marginal cost curves rise from right to left, depicting rising marginal costs of reducing emissions.[4] Exhibit 5.1

[4] In Chapter 3, we showed marginal cost curves sloping upward to the right. The graph goes in the opposite direction because here we are producing *reductions* in emissions.

n with environmental scientists,
develop means by which damage
racy. In the next few chapters we
apter, we will simply use the con-
ential choices that face society in

quantity of residuals being emit-
bient concentrations. Think of the
rse of operation it produces a large
n that it has free access to the river,
is simply to pump them into the
l and managerial means to reduce
se activities are called "abatement
r reducing, the quantity of residu-
on this activity, the pulp mill can
e abatement, the greater the cost.
one source to another, depending
missions of $SO_2$ from electric power
osts of reducing, say, toxic fumes
ucing the same type of effluent the
ecause of differences in the techno-
may be relatively new, using mod-
r may be an old one using more
on that follows keep in mind that
notation and includes all the many
es in production technology, input
andonment of a site, and so forth.

concept of the **marginal abatement**
ne as before: quantities of pollutants
the vertical axis. Marginal emission
ieving a one-unit decrease in emis-
missions are increased by a unit. On
curves originate at the uncontrolled
o undertaking any abatement activ-
ent costs show the marginal costs of
ese marginal cost curves rise from
of reducing emissions.[4] Exhibit 5.1

upward to the right. The graph goes in
*reductions* in emissions.



## The Abatement Cost Function for Cleaning Up Boston Harbor[1]

**EXHIBIT 5.1**

| Cost/ Household/ Year | What You Get | Effects on the Community and the Environment | Legality |
|---|---|---|---|
| $0.00 | No running water; no sewage pipes to remove sewage from houses. | City life impossible; unsafe drinking water leads to disease; local ponds and rivers drained for water; water shortages; sewage in streets causes epidemics; local ponds and rivers destroyed by sewage; major changes in animal life and urban ecology. | ($0) Illegal: Federal CWA and others violated. |
| $125.00 | Running water in your house; clean, safe drinking water; no sewage removed from your house. | City life miserable due to raw sewage in the streets; epidemics caused by raw sewage; rivers, lakes, and harbor polluted with bacteria; destruction of local ponds and rivers by sewage; major changes in animal life and urban ecology; no safe swimming; coastal seafood contaminated. | ($125) Illegal: Federal CWA and others violated. |
| $175.00 | Running water in your house; clean, safe drinking water; sewage piped to harbor—no treatment. | Harbor unswimmable and smelly; health risk presented by raw sewage; harbor polluted by sewage and excess nutrients; shellfish contaminated; no safe ocean swimming; rats feed on fish killed by low oxygen levels. | ($175) Illegal: Federal CWA and others violated. |
| $225.00 | Running water in your house; clean, safe drinking water; sewage removed from house; primary treatment under typical conditions; frequent releases of raw sewage through combined sewer outfalls. | Boston Harbor polluted with bacteria and toxins; health risk presented by raw sewage in harbor; fish growth limited by low oxygen in the summer; all harbor seafood (except lobster) contaminated; beaches closed frequently in summer. | ($225) Illegal: Federal CWA and others violated. |

*(Continued)*

## EXHIBIT 5.1 *(Continued)*

| Cost/ Household/ Year | What You Get | Effects on the Community and the Environment | Legality |
|---|---|---|---|
| $725.00 | Running water in your house; clean, safe drinking water; sewage removed from house; primary treatment under typical conditions; secondary treatment under typical conditions; many releases of raw sewage through CSOs per year. | Improvement in harbor from present; bacterial pollution and low oxygen levels caused by combined sewer outfall (CSO) releases; all harbor seafood (except lobster) contaminated; beaches closed frequently in summer. | ($725) Legal: Under typical conditions. ($725) Illegal: Federal CWA violated during heavy rain storms. |
| $800.00 | Running water in your house; clean, safe drinking water; sewage removed from house; primary treatment; secondary treatment and sludge recycling; long outfall; storage for CSO water; infrequent releases of raw sewage through CSOs. | Improvement in harbor from present; seafood caught in harbor is edible; few or no beaches closed during summer; harbor swimmable under good conditions. | ($800) Legal: Federal CWA requirements met. |
| $1,200.00 | Running water in your house; clean, safe drinking water; sewage removed from house; primary treatment; secondary treatment and sludge recycling; tertiary treatment; long outfall; containment of CSO water. | Sewage has no effect on harbor; healthy marine environment in harbor; harbor swimmable. | ($1,200) Legal: Federal CWA requirements exceeded. |

[1] These abatement costs are in terms of dollars per household per year. They are not, strictly speaking, marginal abatement costs, but you can determine what these are by looking at the differences in costs between the various levels.

**Source:** Exhibit material displayed at the New England Aquarium, Boston, MA, Spring 2000. Thanks to Stephen Costa for finding this material.

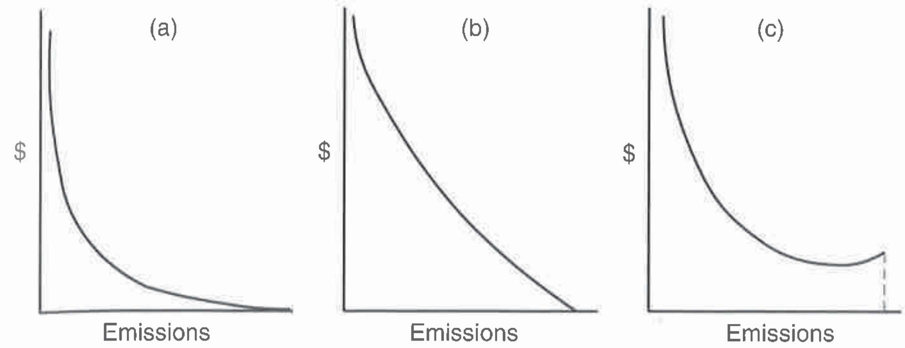
| he and ment | Legality |
|---|---|
| in harbor from erial pollution en levels caused sewer outfall s; all harbor pt lobster) con- aches closed summer. | ($725) Legal: Under typical conditions. ($725) Illegal: Federal CWA violated during heavy rain storms. |
| in harbor from od caught in le; few or no d during sum- wimmable under ons. | ($800) Legal: Federal CWA requirements met. |
| o effect on ny marine in harbor; nable. | ($1,200) Legal: Federal CWA requirements exceeded. |

hold per year. They are not, strictly
ine what these are by looking at the

quarium, Boston, MA, Spring 2000.



**FIGURE 5.3** **Representative Marginal Abatement Cost Functions**



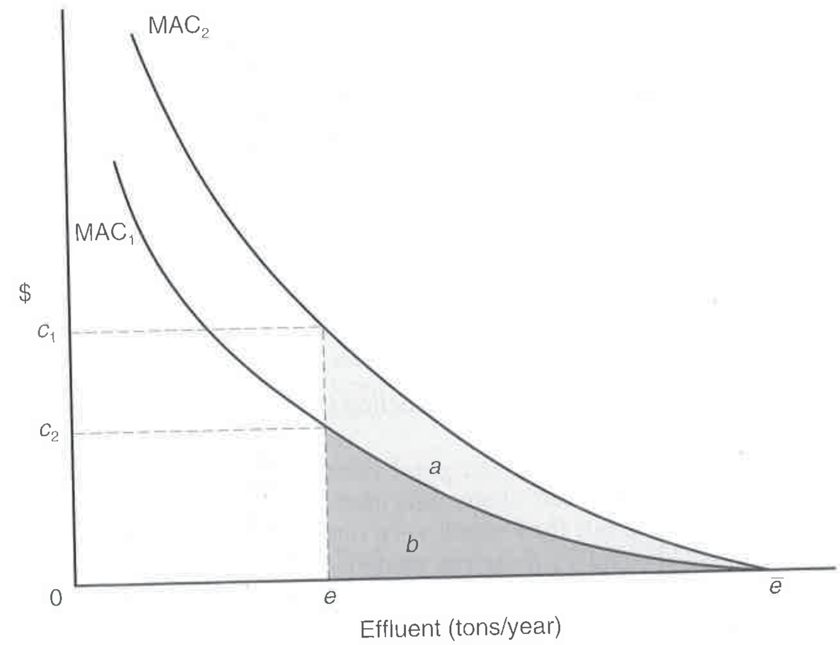

shows data pertaining to the abatement cost function for cleaning up the water of Boston Harbor.

Figure 5.3 shows three alternative marginal abatement cost functions. The one in panel (a) depicts marginal abatement costs rising very modestly as emissions are first reduced, but then rising very rapidly as emissions become relatively small. Panel (b) shows marginal abatement costs that rise rapidly from the beginning. Panel (c) shows a marginal abatement cost curve that has an initial declining phase, followed by increasing values; this might characterize a situation in which small reductions can be handled only by technical means that require a substantial initial investment. For somewhat larger reductions, the marginal costs actually may decline as it becomes possible to utilize these techniques more fully. Ultimately, however, marginal abatement costs increase. We have to keep in mind that in dealing with abatement costs we are dealing with a cost concept similar to that discussed in Chapter 3. The level of costs encountered when carrying out any particular task depends on the technology available to do the task and also on the managerial skills that are applied to the job. It is quite possible to suffer extremely high abatement costs if the wrong technology is used or if what is available is used incorrectly. In other words, the marginal abatement cost functions pictured are to be understood as the **minimum** costs of achieving reductions in emissions.

## Abatement Cost Functions: A Closer Look

To investigate more deeply the concept of marginal abatement cost, consider Figure 5.4, which shows two marginal abatement cost curves. For the moment we focus on the higher one, labeled $MAC_2$. It begins at an effluent level marked $\bar{e}$, the uncontrolled emission level. From there it slopes upward to the left. Beginning at the uncontrolled level, the first units of emission reduction can be achieved with a relatively low marginal cost. Think again of the pulp mill. This first small decrease might be obtained with the addition of a modest settling pond, but as emission levels are reduced further the marginal cost of achieving additional reductions increases. For example, to get a 30–40 percent reduction, the pulp mill may have to invest in new technology that is more efficient in

**FIGURE 5.4 Anatomy of a Marginal Abatement Cost Curve**




terms of water use. A 60–70 percent reduction in effluent might require substantial new treatment technology in addition to all the steps taken previously, whereas a 90–95 percent reduction might take very costly equipment for recycling virtually all of the production residuals produced in the plant. Thus, the larger the reduction in emissions, the greater the marginal costs of producing further reductions. This yields a marginal abatement cost function that gets steeper in slope as emissions are reduced.[5]

Of course, there is an upper limit on these abatement costs. The extreme option for a single plant or pollution source is to cease operations, thereby achieving a zero level of emissions. The costs of doing this depend on circumstances. If the source is just one small plant within a large industry consisting of many such plants, the costs of closing it down may not be that great. In fact it may have very little impact on, say, the price to consumers of whatever is being produced (e.g., paper in the pulp mill), although the local impact on jobs and community welfare may be substantial. But if we are talking about the marginal abatement costs for an entire industry—electric power production in the midwestern United States, for example—the shutdown option, as a way of achieving zero emissions, would have enormous costs.

[5] Remember that the quantity indexed on the horizontal axis is the quantity of emissions, starting at zero on the left. Thus, the marginal abatement costs of reducing emissions increase as you move to the left, that is, as you decrease emissions. In Chapter 3 we introduced marginal cost curves that had the conventional shape of increasing to the right as output increased. If we indexed the *quantity of emissions reduced* starting at zero, then the MAC curve would indeed increase to the right. We think it more intuitive, however, to have the origin correspond to zero actual emissions.

The marginal abatement cost function can express **actual** marginal costs of a source or group of sources or the **lowest possible** marginal abatement costs. Actual costs, of course, are determined by the technologies and procedures that firms have adopted in the past to reduce emissions. These could have been affected by a variety of factors, including managerial shortsightedness or public pollution control regulations. To use the model for studying questions of social efficiency and cost effectiveness, however, we don't want actual costs but the lowest possible abatement costs. In this case, we have to assume that sources have adopted whatever technological and managerial means are available to achieve emission reductions at the lowest possible costs. We have to assume, in other words, that sources are acting in a **cost-effective** manner.

As with any marginal graph, we can depict not only marginal but also total values. If emissions are currently at $e$ tons per year, the value on the vertical axis shows the marginal cost of achieving one more unit of emission reduction. The area under the marginal abatement cost curve, between its origin at point $\bar{e}$ and any particular emission level, is equal to the **total costs** of abating emissions to that level. For example, with the curve labeled $MAC_2$, the total abatement cost of achieving an emission level of $e$ tons per year is equal to the area under the curve between $e$ and $\bar{e}$, the area $(a + b)$; remember that we are reading the graph from right to left.

Consider now the other marginal abatement cost curve shown in Figure 5.4, labeled $MAC_1$. Its main feature is that it lies below $MAC_2$, meaning that it corresponds to a situation where the marginal abatement costs for any level of emissions are lower than those of $MAC_2$. At $e$ tons per year of emissions, for example, the marginal costs of abating an extra ton are only $c_2$ in the case of $MAC_1$, which are substantially lower than the marginal abatement costs of $MAC_2$ at this point. What could account for the difference? Let us assume that we are dealing with the same pollutant in each case. One possibility is that these apply to different sources—for example, a plant that was built many years ago and another that was built more recently and uses different production technology. The newer plant lends itself to less costly emissions reduction.

Another possibility is that $MAC_1$ and $MAC_2$ relate to the same pollutant and the same source, but at different times. The lower one represents the situation after a new pollution-control technology has been developed, whereas the upper one applies before the change. Technological change, in other words, results in a lowering of the marginal abatement cost curve for a given pollutant. It is possible to represent graphically the annual cost that this source would save assuming the emission rate is $e$ before and after the change. Before the firm adopted the new technology, its total abatement cost of achieving effluent level $e$ was equal to $(a + b)$ per year, whereas after the change the total abatement costs are $b$ per year. The annual cost savings from the technological change are thus $a$. This type of analysis will be important when we examine different types of pollution-control policies because one of the criteria we will want to use to evaluate these policies is how much cost-saving incentive they offer to firms to engage in research and development to produce new pollution-control technologies.

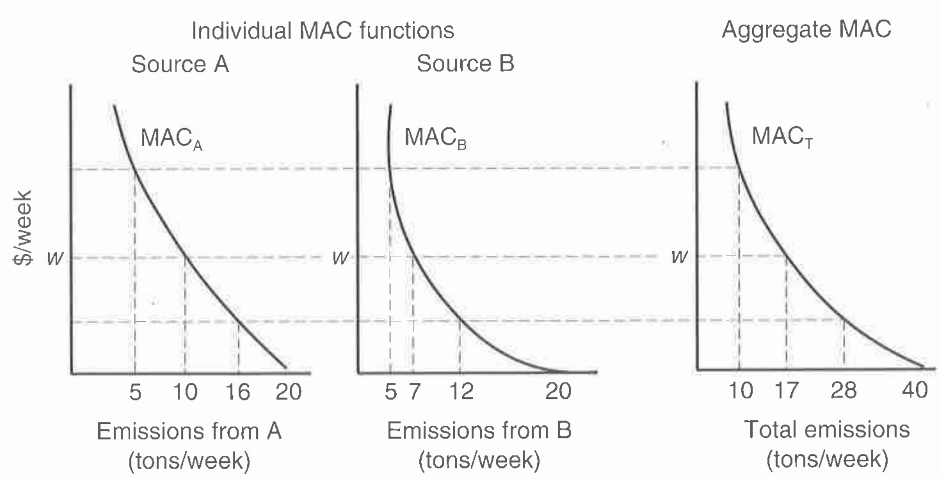
## Aggregate Marginal Abatement Costs

The discussion of the last few pages has treated the marginal abatement cost function as something applying to a single firm, such as a single pulp mill on a river. Suppose, however, we want to talk about the marginal abatement cost of a group of firms, perhaps a group of firms in the same industry or a group of firms all located in the same region. Most environmental policies, especially at state or federal levels, are aimed at controlling emissions from groups of pollution sources, not just single polluters. Suppose, furthermore, that the individual marginal abatement cost functions differ among the various firms. To control organic pollutants in Boston Harbor or San Francisco Bay, for example, would require controlling emissions from a large variety of different sources in different industries with different production technologies, and therefore with very different individual marginal abatement cost functions. In this case we would have to construct the overall, or **aggregate, marginal abatement cost function** for the collection of firms by adding together the individual marginal abatement cost curves.

Although this sounds simple, and it basically is, it nevertheless leads into one of the more important concepts underlying the design of effective environmental policy. It is critical to keep in mind the central idea of the abatement cost function. It is a function that shows the *least costly* way of achieving reductions in emissions for an individual firm if we are looking at an individual marginal abatement cost function, or for a group of polluting sources if we are considering the aggregate marginal abatement cost function.

Figure 5.5 shows, on the left, two individual marginal abatement cost functions, labeled Source A and Source B. Note that they are not the same (although remember that the scales are the same; that is, we are dealing with the same pollutant). $MAC_A$ starts at 20 tons/week and rises rather rapidly as emissions are reduced. $MAC_B$ also begins at the uncontrolled discharge level of 20 tons/week,

**FIGURE 5.5** Aggregate Abatement Costs

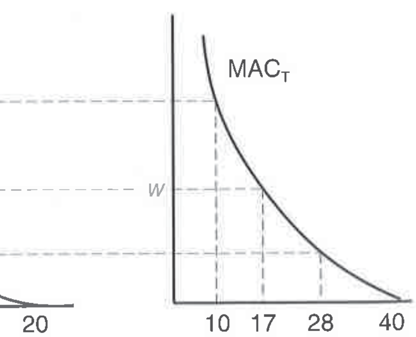


## osts

eated the marginal abatement cost
rm, such as a single pulp mill on a
out the marginal abatement cost of
in the same industry or a group
environmental policies, especially
trolling emissions from groups of
uppose, furthermore, that the indi-
differ among the various firms. To
or San Francisco Bay, for example,
large variety of different sources in
n technologies, and therefore with
ent cost functions. In this case we
**gregate, marginal abatement cost**
g together the individual marginal

ally is, it nevertheless leads into one
e design of effective environmental
al idea of the abatement cost func-
*tly* way of achieving reductions in
looking at an individual marginal
olluting sources if we are consider-
unction.

ual marginal abatement cost func-
hat they are not the same (although
, we are dealing with the same pol-
ises rather rapidly as emissions are
ed discharge level of 20 tons/week,






but rises much less rapidly. Why the difference? Perhaps Source B is a newer plant with more flexible technological alternatives for pollution control. Or perhaps the two sources, although producing the same type of effluent, are manufacturing different consumer goods and using different production techniques. For whatever reason, they have different marginal abatement cost curves.

The aggregate marginal abatement cost curve is a summation, or aggregation, of these two individual relationships. But since the individual curves are different, it makes a great deal of difference how they are added together. The problem is that when there are two (or any other number greater than one) sources with different abatement costs, the total cost will depend on how the total emissions are allocated among the different sources. The principle to follow is to add together the two individual functions in such a way as to yield the lowest possible aggregate marginal abatement costs. The way to do this is to add them horizontally. Select a particular level of marginal abatement cost—for example, the one marked $w$ in Figure 5.5. This level of marginal abatement cost is associated with an effluent level of 10 tons/week from Source A and an effluent level of about 7 tons/week from Source B. On the aggregate curve, thus, a marginal abatement cost of $w$ would be associated with an effluent level of 10 tons + 7 tons = 17 tons/week. All the other points on the aggregate marginal abatement cost curve are found the same way, by summing across horizontally on the individual marginal abatement cost curves.

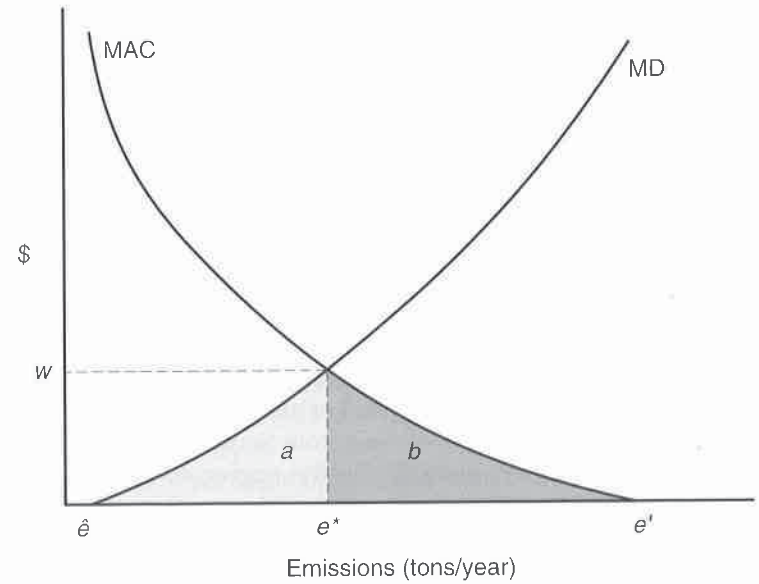
In effect, what we have done here is to invoke the important **equimarginal principle,** an idea that was introduced earlier in Chapter 4. To get the minimum aggregate marginal abatement cost curve, the aggregate level of emissions must be distributed among the different sources in such a way that they all have the same marginal abatement costs. Start at the 10 tons/week point on the aggregate curve. Obviously, this 10-ton total could be distributed among the two sources in any number of ways: 5 tons from each source, 8 tons from one and 2 from the other, and so on. Only one allocation, however, will give the lowest aggregate marginal abatement costs; this is the allocation that leads the different sources to the point at which they have exactly the same marginal abatement costs. At the end of this chapter we will come back to this equimarginal principle, illustrating it with a simple numerical example.

## The Socially Efficient Level of Emissions

We have considered separately the marginal damage function and the marginal abatement cost function related to a particular pollutant being released at a particular place and time; it is now time to bring these two relationships together. This we do in Figure 5.6, which depicts a set of conventionally shaped marginal damage and marginal abatement cost curves labeled, respectively, MD and MAC. Marginal damages have a threshold at emission level $\hat{e}$, whereas the uncontrolled emission level is $e'$.

The efficient level of emissions is defined as that level at which marginal damages are equal to marginal abatement costs. What is the justification for this?

**FIGURE 5.6** The Efficient Level of Emissions




Note the trade-off that is inherent in the pollution phenomenon: higher emissions expose society, or some part of it, to greater costs stemming from environmental damages. Lower emissions involve society in greater costs in the form of resources devoted to abatement activities. The efficient level of emissions is thus the level at which these two types of costs exactly offset one another—that is, where **marginal abatement costs equal marginal damage costs.** This is emission level $e^*$ in Figure 5.6. Marginal damages and marginal abatement costs are equal to each other and to the value $w$ at this level of emissions.

We also can look at this outcome in terms of total values because we know that the totals are the areas under the marginal curves. Thus, the triangular area marked $a$ (bounded by points $\hat{e}$ and $e^*$ and the marginal damage function) depicts the total damages existing when emissions are at level $e^*$, whereas the triangular area $b$ shows the total abatement costs at this level of emissions. The sum of these two areas $(a + b)$ is a measure of the total social costs from $e^*$ tons per year of this particular pollutant. The point $e^*$ is the unique point at which this sum is minimized. Note that the size of area $a$ need not equal the size of area $b$.

You might get the impression, on the basis of where point $e^*$ is located on the $x$-axis, that this analysis has led us to the conclusion that the efficient level of emissions is always one that involves a relatively large quantity of emissions and substantial environmental damages. This is not the case. What we are developing, rather, is a conceptual way of looking at a trade-off. In the real world every pollution problem is different. This analysis gives us a generalized way of framing the problem that obviously has to be adapted to the specifics of any particular case of environmental pollution. Figure 5.7, for example, depicts three different situations that might characterize particular environmental sions

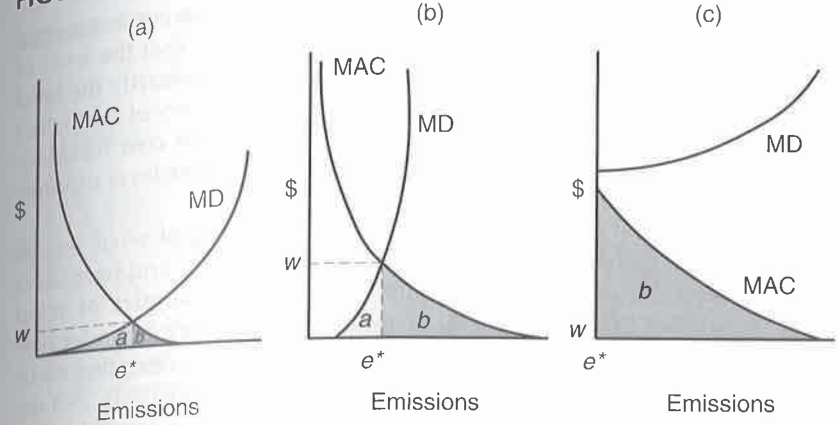

e pollution phenomenon: higher emis-
o greater costs stemming from environ-
lve society in greater costs in the form of
es. The efficient level of emissions is thus
osts exactly offset one another—that is,
**marginal damage costs.** This is emission
and marginal abatement costs are equal
evel of emissions.

rms of total values because we know that
ginal curves. Thus, the triangular area
d the marginal damage function) depicts
ons are at level $e^*$, whereas the triangular
this level of emissions. The sum of these
l social costs from $e^*$ tons per year of this
unique point at which this sum is mini-
not equal the size of area *b*.

basis of where point $e^*$ is located on the
the conclusion that the efficient level of
a relatively large quantity of emissions
es. This is not the case. What we are de-
f looking at a trade-off. In the real world
is analysis gives us a generalized way of
as to be adapted to the specifics of any
ution. Figure 5.7, for example, depicts
characterize particular environmental



**FIGURE 5.7** **Efficient Emission Levels for Different Pollutants**




pollutants. In each case $e^*$ depicts the efficient level of emissions and *w* shows marginal damages and marginal abatement costs at that quantity of emissions. Panel (a) shows a pollutant for which $e^*$ is well to the right of zero (of course, since the horizontal axis has no units, it's not clear exactly what "well to the right" actually means here). Marginal damages at this point are quite small; so are total damages and abatement costs, as shown by the small size of the triangles corresponding to these values. The reason is that this is a pollutant where both marginal abatement costs and marginal damages increase at first only very slowly.

Panel (b) shows a situation where the marginal abatement function rises moderately, then rapidly, whereas the marginal damage function rises very rapidly from the beginning. In this case $e^*$ is well to the right of zero, and *w* lies well above what it was in the first diagram (assuming the vertical axes of these diagrams all have the same scale). Note, however, that at $e^*$ total abatement costs are substantially higher than total damages, as is indicated by the relative sizes of the triangles that measure these total values (*a* and *b*). What this emphasizes is that it is not the equality of total abatement costs and total damages that defines the efficient level of effluent, but the equality of the **marginal abatement costs** and **marginal damages.**

In panel (c) of Figure 5.7 the efficient level of emissions is zero. There is no point of intersection of the two functions in the graph; area *a* does not even appear on the graph. The only way we could conceivably get them to intersect is if we could somehow extend them to the left of the vertical axis, but this would imply that emissions could actually be negative, which is an oddity that we will avoid. What makes $e^* = 0$ is that the marginal damage function doesn't begin at zero, but rather well up on the *y*-axis, implying that even the first small amount of this pollutant placed in the environment causes great damage (perhaps this diagram applies to some extremely toxic material). Relative to this the marginal costs of abatement are low, giving an efficient emission level of zero.

## Changes in the Efficient Level of Emissions

The real world is a dynamic place, and this is especially true of environmental pollution control. For our purposes this implies, for example, that the level of emissions that was efficient last year, or last decade, is not necessarily the level that is efficient today or that is likely to be in the future. When any of the factors that lie behind the marginal damage and marginal abatement cost functions change, the functions themselves will shift and $e^*$, the efficient level of emissions, also will change.

Before taking a look at this, we need to remind ourselves of what we are doing. Remember the distinction made earlier between positive and normative economics, between the **economics of *what is*** and the **economics of *what ought to be*.** The idea of the efficient level of emissions comes firmly under normative economics, under the idea of what ought to be. We are presenting emission level $e^*$, the level that balances abatement costs and damage costs, as a desirable target for public policy. Do not get this confused with the actual level of emissions. If the world worked so that the actual level of emissions was always equal to, or close to, the efficient level, we presumably would have no need to worry about intervening with environmental policy of one type or another. Of course it does not, which is why we must turn to public policy.

Figure 5.8 shows several ways in which $e^*$ might change when underlying factors change. Panel (a) shows the results of a shift upward in the marginal damage function, from $MD_1$ to $MD_2$. One of the ways this could happen is through population growth. $MD_1$ might apply to a municipality in 1980 and $MD_2$ to the same municipality in 2000 after its population has grown. More people means that a given amount of effluent will cause more damage.[6] This leads to a conclusion that is intuitively straightforward: The efficient level of emissions drops from $e^*_1$ to $e^*_2$. With a higher marginal damage function, the logic of the efficiency trade-off would lead us to devote more resources to pollution control.

Panel (b) of Figure 5.8 shows the case of a shift in the marginal abatement cost function, from $MAC_1$ to $MAC_2$. What could have caused this? The most obvious, perhaps, is a change in the technology of pollution control. As stressed earlier, abatement costs depend critically on the technology available for reducing effluent streams: treatment technology, recycling technology, alternative fuel technology, and so forth. New techniques normally arise because resources, talents, and energy have been devoted to research and development. So the shift downward in marginal abatement costs depicted in Figure 5.8 might be the result of the development of new treatment or recycling technologies that make it less costly to reduce the effluent stream of this particular pollutant. It should not be too surprising that this leads to a reduction in the efficient level of emissions, as indicated by the change from $e^*_1$ to $e^*_2$. We might note that this could lead to either an increase or a decrease in the total cost of abating

[6] This diagram also could apply, of course, to a different situation. $MD_1$ could be the damage function pertaining to a relatively sparsely settled rural region; $MD_2$ could be the marginal damage function pertaining to a more-populous urban area. Everything we say about the relationship between $e^*_1$ and $e^*_2$ applies also to cases like this where we are comparing two different places at the same time, in addition to the above comparison of the same place at two different times.

## nissions

is especially true of environmental plies, for example, that the level of decade, is not necessarily the level the future. When any of the factors narginal abatement cost functions and $e^*$, the efficient level of emis-

remind ourselves of what we are er between positive and normative ***at is*** and the **economics of *what*** emissions comes firmly under nor- ught to be. We are presenting emis- nt costs and damage costs, as a de- is confused with the actual level of ctual level of emissions was always resumably would have no need to al policy of one type or another. Of rn to public policy.

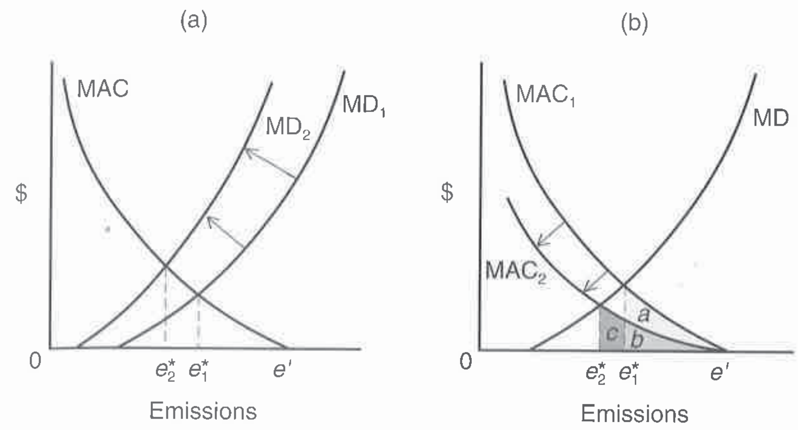
might change when underlying fac- shift upward in the marginal dam- ways this could happen is through unicipality in 1980 and $MD_2$ to the on has grown. More people means ore damage.[6] This leads to a conclu- e efficient level of emissions drops nage function, the logic of the effi- re resources to pollution control.

f a shift in the marginal abatement uld have caused this? The most ob- gy of pollution control. As stressed the technology available for reduc- , recycling technology, alternative es normally arise because resources, research and development. So the sts depicted in Figure 5.8 might be ment or recycling technologies that stream of this particular pollutant. eads to a reduction in the efficient e from $e^*_1$ to $e^*_2$. We might note that lecrease in the total cost of abating

situation. $MD_1$ could be the damage region; $MD_2$ could be the marginal damage erything we say about the relationship we are comparing two different places at the same place at two different times.



**FIGURE 5.8 Changes in $e^*$, the Efficient Level of Emissions**




emissions. Before the change, total abatement costs were an amount equal to the area $(a + b)$, that is, the area under $MAC_1$ between the uncontrolled level $e'$ and the amount $e^*_1$. After the change, total abatement costs are equal to area $(b + c)$, and the question of whether total abatement costs at the efficient level of emissions have increased or decreased hinges on the relative sizes of the two areas $a$ and $c$. This in turn depends on the shapes of the curves and the extent to which the marginal abatement cost curve has shifted; the more it has shifted, the more likely it is that the efficient level of total abatement costs after the change will exceed the costs before the change.[7]

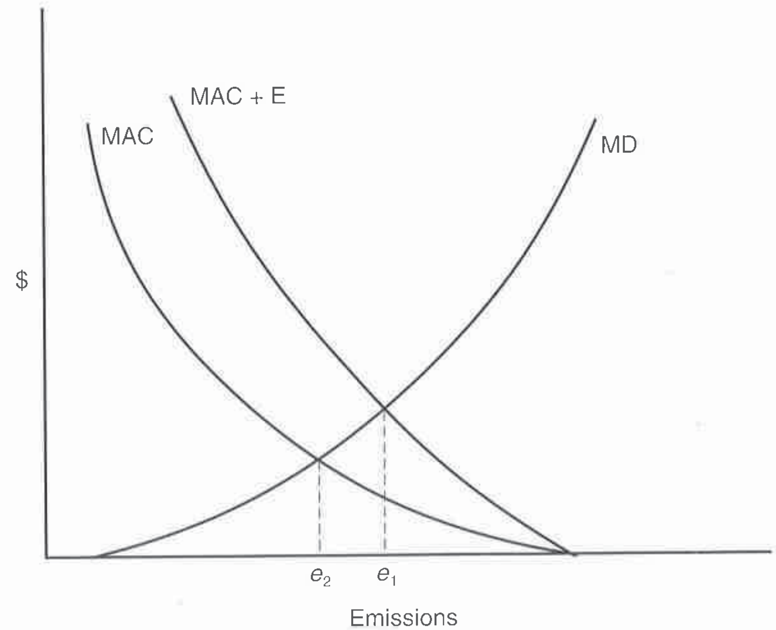
## Enforcement Costs

So far the analysis has considered only the private costs of reducing emissions, but emission reductions do not happen unless resources are devoted to enforcement. To include all sources of cost we need to add **enforcement costs** to the analysis. Some of these are private, such as added recordkeeping by polluters, but the bulk are public costs related to various regulatory aspects of the enforcement process.

Figure 5.9 shows a simple model of pollution control with enforcement costs added. To the normal marginal abatement cost function has been added the marginal costs of enforcement, giving a total marginal cost function labeled MAC + E. The vertical distance between the two marginal cost curves equals marginal enforcement costs. The assumption drawn into the graph is that marginal enforcement costs, the added costs of enforcement that it takes to get emissions reduced by a unit, increase as emissions decrease. In other words, the more

[7] These diagrams also can be used to examine some of the implications of making mistakes. For example, suppose the public control authorities think that the real marginal abatement cost was $MAC_1$, but, in fact, because there is a cheaper way of reducing this effluent that they do not know about, marginal abatement costs are actually $MAC_2$. Then we would conclude that the efficient level of effluent is $e^*_1$, whereas it is actually $e^*_2$. We might be shooting at a target that involves excessive emissions.

**FIGURE 5.9** **Enforcement Costs**




polluters cut back emissions, the more costly it is to enforce further cutbacks. We will have more to say about enforcement and its costs in later parts of the book.

In effect, the addition of enforcement costs moves the efficient level of emissions to the right of where it would be if they were zero. This shows the vital importance of having good enforcement technology because lower marginal enforcement costs would move MAC + E closer to MAC, decreasing the efficient emission level. In fact, **technical change in enforcement** has exactly the same effect on the efficient level of emissions as technical change in emissions abatement. We will have more to say about enforcement in later chapters, especially Chapter 11.

## The Equimarginal Principle Applied to Emission Reductions

Before going on, we will take a last, very explicit look at the equimarginal principle. In the present context, the application of the equimarginal principle says the following: If there are **multiple sources** of a particular type of pollutant with **differing marginal abatement costs,** and if it is desired to reduce aggregate emissions at the **least possible cost** (or alternatively, get the greatest reduction in emissions for a given cost), then emissions from the various sources must be reduced in accordance with the **equimarginal principle.**

To illustrate this, look at the numbers in Table 5.2. This shows explicitly the marginal abatement costs of each of two firms emitting a particular residual into the environment. If neither source makes any effort to control emissions, they will


it is to enforce further cutbacks. We
l its costs in later parts of the book.
moves the efficient level of emissions
ero. This shows the vital importance
cause lower marginal enforcement
C, decreasing the efficient emission
**nt** has exactly the same effect on the
ge in emissions abatement. We will
hapters, especially Chapter 11.

## Applied

plicit look at the equimarginal prin-
of the equimarginal principle says
f a particular type of pollutant with
f it is desired to reduce aggregate
rnatively, get the greatest reduction
s from the various sources must be
**al principle.**
Table 5.2. This shows explicitly the
s emitting a particular residual into
effort to control emissions, they will



**TABLE 5.2 The Equimarginal Principle**

| | Marginal Abatement Costs ($1,000/week) | |
|---|---|---|
| Emissions (tons/week) | Source A | Source B |
| 12 | 0 | 0 |
| 11 | 1 | 2 |
| 10 | 2 | 4 |
| 9 | 3 | 6 |
| 8 | 4 | 10 |
| 7 | 5 | 14 |
| 6 | 6 | 20 |
| 5 | 8 | 25 |
| 4 | 10 | 31 |
| 3 | 14 | 38 |
| 2 | 24 | 58 |
| 1 | 38 | 94 |
| 0 | 70 | 160 |

each emit 12 tons/week. If Plant A reduces its emissions by 1 ton, to 11 tons/week, it will cost $1,000/week; if it reduces effluent further to 10 tons/week, its abatement costs will increase by $2,000/week, and so on. Note that the marginal abatement cost relationships of the two sources are different: that of Source B increases faster than that of Source A.

Suppose that initially each plant is emitting at the uncontrolled level; total emissions would then be 24 tons/week. Now assume that we want to reduce overall emissions to half the present level, or a total of 12 tons/week. One way to do this would be to have **equiproportionate** cutbacks. Because we want a total reduction of 50 percent, each source is required to reduce by 50 percent. If Source A were cut 50 percent to 6 tons/week, its marginal abatement costs at this level would be $6,000/week, whereas at this level of emissions the marginal abatement costs of Source B would be $20,000/week. Total abatement costs of the 12-ton total can be found by adding up the marginal abatement costs; these are $21,000/week for Source A ($1,000 + $2,000 + $3,000 + $4,000 + $5,000 + $6,000) and $56,000/week for Source B ($2,000 + $4,000 + $6,000 + $10,000 + $14,000 + $20,000), or a grand total of $77,000/week.

The overall reduction to 12 tons/week, however, can be achieved with a substantially lower total cost. We know this because the equiproportionate reduction violates the equimarginal principle; marginal abatement costs are not equalized when each source reduces its effluent to 6 tons/week. What is required is different emission rates for the two sources, where, simultaneously, they will emit no more than 12 tons of effluent and have the same marginal abatement costs. This condition is satisfied if Source A emits 4 tons and Source B emits 8 tons. These rates add up to 12 tons total and give each source a marginal abatement cost of $10,000/week. Calculating total abatement costs at these emission levels gives $39,000/week for Source A ($1,000 + $2,000 + $3,000 + $4,000 + $5,000 + $6,000 + $8,000 + $10,000) plus $22,000/week for Source B

($2,000 + $4,000 + $6,000 + $10,000), or a grand total of $61,000/week. By following the equimarginal principle, the desired reduction in total emissions has been obtained, but with a savings of $16,000/week over the case of an equiproportionate reduction.

Thus, we see that an emission reduction plan that follows the equimarginal rule gives emission reduction at minimum cost. Another way of saying this is that for any particular amount of money devoted to effluent reduction, the maximum quantitative reduction in total effluent can only be obtained by following the equimarginal principle. The importance of this principle cannot be overstated. When defining the efficient level of emissions, we were going on the assumption that we were working with the lowest possible marginal abatement cost function. The only way of achieving this is by controlling individual sources in accordance with the equimarginal rule. If we are designing public policy under the rule of equiproportionate reductions at the various sources, the marginal abatement cost function will be higher than it should be. One of the results of this is that the efficient emission level will be higher than it should be, or, to say the same thing, we will seek smaller reductions in emissions than are socially efficient.

## Trade-offs and Politics

The basic model presented in the chapter says that, in deciding how to manage the level of environmental pollution, society faces a trade-off between damage reduction and abatement costs. In the rough-and-tumble of the policy process not everybody is happy with this idea. It is a political fact of life that different social groups are more heavily invested in one side than the other. The environmental community puts relatively great stress on the benefits of pollution control (damage reduction); the regulated community emphasizes the cost side of the trade-off. Neither side may be particularly happy with an approach that gives equal weight to each side. Both sides will find much to object to in how the opposite side is assessing impacts, especially when there is a lot of uncertainty surrounding both abatement cost and damage functions.

## Summary

In this chapter we have looked at a simple model of pollution control. It is based on the notion of a trade-off between environmental damages and pollution abatement costs. We introduced the notion of a **marginal damage function,** showing the marginal social damages resulting from varying levels of residual emissions or ambient pollutant levels. Then we looked at **marginal abatement cost** relationships, first for an individual pollution source and then for a group of such sources. By bringing together these two types of relationships we then defined an **efficient level of emissions:** that level at which marginal damages and marginal abatement costs are equal. At this level of emissions total social costs, the total of abatement costs and damages, are minimized.