2 A Brief Review of Microeconomic Theory

Practical men, who believe themselves to be quite exempt from any intellectual influences, are usually the slaves of some defunct economist. . . . It is ideas, not vested interests, which are dangerous for good or evil.

JOHN MAYNARD KEYNES,
THE GENERAL THEORY OF EMPLOYMENT, INTEREST, AND MONEY (1936)

In this state of imbecility, I had, for amusement, turned my attention to political economy.

THOMAS DEQUINCEY,
CONFESSIONS OF AN ENGLISH OPIUM EATER (1821)

Economics is the science which studies human behavior as a relationship between ends and scarce means which have alternative uses.

LIONEL CHARLES ROBBINS, LORD ROBBINS,
AN ESSAY ON THE NATURE AND SIGNIFICANCE OF ECONOMIC SCIENCE (1932)

THE ECONOMIC ANALYSIS of law draws upon the principles of microeconomic theory, which we review in this chapter. For those who have not studied this branch of economics, reading this chapter will prove challenging but useful for understanding the remainder of the book. For those who have already mastered microeconomic theory, reading this chapter is unnecessary. For those readers who are somewhere in between these extremes, we suggest that you begin reading this chapter, skimming what is familiar and studying carefully what is unfamiliar. If you're not sure where you lie on this spectrum of knowledge, turn to the questions at the end of the chapter. If you have difficulty answering them, you will benefit from studying this chapter carefully.

I. Overview: The Structure of Microeconomic Theory

Microeconomics concerns decision making by individuals and small groups, such as families, clubs, firms, and governmental agencies. As the famous quote from Lord Robbins at the beginning of the chapter says, microeconomics is the study of how scarce resources are allocated among competing ends. Should you buy that digital audiotape player you'd like, or should you buy a dapper suit for your job interview? Should you take a trip with some friends this weekend or study at home? Because you have limited income and time and cannot, therefore, buy or do everything that you might want to buy or do, you have to make choices. Microeconomic theory offers a general theory about how people make such decisions.

We divide our study of microeconomics into five sections. The first is the theory of consumer choice and demand. This theory describes how the typical consumer, constrained by a limited income, chooses among the many goods and services offered for sale.

The second section deals with the choices made by business organizations or firms. We shall develop a model of the firm that helps us to see how the firm decides what goods and services to produce, how much to produce, and at what price to sell its output. In the third section, we shall consider how consumers and firms interact. By combining the theory of the consumer and the firm, we shall explain how the decisions of consumers and firms are coordinated through movements in market price. Eventually, the decisions of consumers and firms must be made consistent in the sense that somehow the two sides agree about the quantity and price of the good or service that will be produced and consumed. When these consumption and production decisions are consistent in this sense, we say that the market is in equilibrium. We shall see that powerful forces propel markets toward equilibrium, so that attempts to divert the market from its path are frequently ineffectual or harmful.

The fourth section of microeconomic theory describes the supply and demand for inputs into the productive process. These inputs include labor, capital, land, and managerial talent; more generally, inputs are all the things that firms must acquire in order to produce the goods and services that consumers or other firms wish to purchase.

The final section of microeconomics deals with the area known as welfare economics. There we shall discuss the organization of markets and how they achieve efficiency.

These topics constitute the core of our review of microeconomic theory. There are four additional topics that do not fit neatly into the sections noted above but that we think you should know about them in order to understand the economic analysis of legal rules and institutions. These are game theory, the economic theory of decision making under uncertainty, growth theory, and behavioral economics. We shall cover these four topics in the final sections of this chapter.

II. Some Fundamental Concepts: Maximization, Equilibrium, and Efficiency

Economists usually assume that each economic actor maximizes something: Consumers maximize utility (that is, happiness or satisfaction), firms maximize profits, politicians maximize votes, bureaucracies maximize revenues, charities maximize social welfare, and so forth. Economists often say that models assuming maximizing behavior work because most people are rational, and rationality requires maximization.

One conception of rationality holds that a rational actor can rank alternatives according to the extent that they give her what she wants. In practice, the alternatives available to the actor are constrained. For example, a rational consumer can rank alternative bundles of consumer goods, and the consumer's budget constrains her choice among them. A rational consumer should choose the best alternative that the constraints allow. Another common way of understanding this conception of rational behavior is to recognize that consumers choose alternatives that are well suited to achieving their ends.

Choosing the best alternative that the constraints allow can be described mathematically as *maximizing*. To see why, consider that the real numbers can be ranked from small to large, just as the rational consumer ranks alternatives according to the extent that they give her what she wants. Consequently, better alternatives can be associated with larger numbers. Economists call this association a "utility function," about which we shall say more in the following sections. Furthermore, the constraint on choice can usually be expressed mathematically as a "feasibility constraint." Choosing the best alternative that the constraints allow corresponds to maximizing the utility function subject to the feasibility constraint. So, the consumer who goes shopping is said to maximize utility subject to her budget constraint.

Turning to the second fundamental concept, there is no habit of thought so deeply ingrained among economists as the urge to characterize each social phenomenon as an *equilibrium* in the interaction of maximizing actors. An equilibrium is a pattern of interaction that persists unless disturbed by outside forces. Economists usually assume that interactions tend toward an equilibrium, regardless of whether they occur in markets, elections, clubs, games, teams, corporations, or marriages.

There is a vital connection between maximization and equilibrium in microeconomic theory. We characterize the behavior of every individual or group as maximizing something. Maximizing behavior tends to push these individuals and groups toward a point of rest, an equilibrium. They certainly do not intend for an equilibrium to result; instead, they simply try to maximize whatever it is that interests them. Nonetheless, the interaction of maximizing agents usually results in an equilibrium.

A *stable* equilibrium is one that will not change unless outside forces intervene. To illustrate, the snowpack in a mountain valley is in stable equilibrium, whereas the snowpack on the mountain's peak may be in unstable equilibrium. An interaction headed toward a stable equilibrium actually reaches this destination unless outside forces divert it. In social life, outside forces often intervene before an interaction reaches equilibrium. Nevertheless, equilibrium analysis makes sense. Advanced microeconomic theories of growth, cycles, and disequilibria exist, but we shall not need them in this book. The comparison of equilibria, called comparative statics, will be our basic approach.

Turning to the third fundamental concept, economists have several distinct definitions of *efficiency*. A production process is said to be productively efficient if either of two conditions holds:

1. It is not possible to produce the same amount of output using a lower-cost combination of inputs, or
2. It is not possible to produce more output using the same combination of inputs.

Product

Consider a firm that uses labor and machinery to produce a consumer good called a "widget." Suppose that the firm currently produces 100 widgets per week using 10 workers and 15 machines. The firm is productively efficient if

1. it is not possible to produce 100 widgets per week by using 10 workers and fewer than 15 machines, or by using 15 machines and fewer than 10 workers, or
2. it is not possible to produce more than 100 widgets per week from the combination of 10 workers and 15 machines.

The other kind of efficiency, called *Pareto efficiency* after its inventor[1] or sometimes referred to as *allocative efficiency*, concerns the satisfaction of individual preferences. A particular situation is said to be *Pareto* or *allocatively efficient* if it is impossible to change it so as to make at least one person better off (in his own estimation) without making another person worse off (again, in his own estimation). For simplicity's sake, assume that there are only two consumers, Smith and Jones, and two goods, umbrellas and bread. Initially, the goods are distributed between them. Is the allocation Pareto efficient? Yes, if it is impossible to reallocate the bread and umbrellas so as to make either Smith or Jones better off without making the other person worse off.[2]

These three basic concepts—maximization, equilibrium, and efficiency—are fundamental to explaining economic behavior, especially in decentralized institutions like markets that involve the coordinated interaction of many different people.

III. Mathematical Tools

You may have been anxious about the amount of mathematics that you will find in this book. There is not much. We use simple algebra and graphs.

A. Functions

Economics is rife with functions: production functions, utility functions, cost functions, social welfare functions, and others. A *function* is a relationship between two sets of numbers such that for each number in one set, there corresponds exactly one number in the other set. To illustrate, the columns below correspond to a functional relationship between the numbers in the left-hand column and those in the right-hand column. Thus, the number 4 in the x-column below corresponds to the number 10 in the y-column.

In fact, notice that each number in the x-column corresponds to exactly one number in the y-column. Thus, we can say that the variable y is a function of the variable x, or in the most common form of notation.

$$y = f(x).$$

[1] Vilfredo Pareto was an Italian-Swiss political scientist, lawyer, and economist who wrote around 1900.

[2] There is another efficiency concept—a potential Pareto improvement or Kaldor-Hicks efficiency—that we describe in section IX.C that follows.

This is read as "y is a function of x" or "y equals some f of x."

y-column	*x*-column
2	3
3	0
10	4
10	6
12	9
7	12

Note that the number 4 is not the only number in the x-column that corresponds to the number 10 in the y-column; the number 6 also corresponds to the number 10. In this table, for a given value of x, there corresponds one value of y, but for some values of y, there corresponds more than one value of x. A value of x determines an exact value of y, whereas a value of y does not determine an exact value of x. Thus, in $y = f(x)$, y is called the *dependent variable*, because it depends on the value of x, and x is called the *independent variable*. Because y depends upon x in this table, y is a function of x, but because x does not (to our knowledge) depend for its values on y, x is not a function of y.

Now suppose that there is another dependent variable, named z, that also depends upon x. The function relating z to x might be named g:

$$z = g(x).$$

When there are two functions, $g(x)$ and $f(x)$, with different dependent variables, z and y, remembering which function goes with which variable can be hard. To avoid this difficulty, the same name is often given to a function and the variable determined by it. Following this strategy, the preceding functions would be renamed as follows:

$$y = f(x) \Rightarrow y = y(x),$$
$$z = g(x) \Rightarrow z = z(x).$$

Sometimes an abstract function will be discussed without ever specifying the exact numbers that belong to it. For example, the reader might be told that y is a function of x, and never be told exactly which values of y correspond to which values of x. The point then is simply to make the general statement that y depends upon x but in an as yet unspecified way. If exact numbers are given, they may be listed in a table, as we have seen. Another way of showing the relationship between a dependent and an independent variable is to give an exact equation. For example, a function $z = z(x)$ might be given the exact form

$$z = z(x) = 5 + x/2,$$

which states that the function z matches values of x with values of z equal to five plus one-half of whatever value x takes. The table below gives the values of z associated with several different values of x:

z-column	x-column
6.5	3
12.5	15
8.0	6
6.0	2
9.5	9

A function can relate a dependent variable (there is always just one of them to a function) to more than one independent variable. If we write $y = h(x, z)$, we are saying that the function h matches one value of the dependent variable y to every pair of values of the independent variables x and z. This function might have the specific form

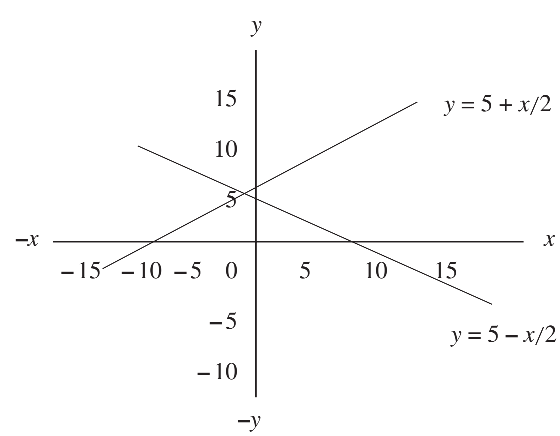
$$y = h(x, z) = -3x + z,$$

according to which y decreases by 3 units when x increases by 1 unit, and y increases by 1 unit when z increases by 1 unit.

B. Graphs

We can improve the intuitive understanding of a functional relationship by visualizing it in a graph. In a graph, values of the independent variable are usually read off the horizontal axis, and values of the dependent variable are usually read off the vertical axis. Each point in the grid of lines corresponds to a pair of values for the variables. For an example, see Figure 2.1. The upward-sloping line on the graph represents all of the pairs of values that satisfy the function $y = 5 + x/2$. You can check this by finding a couple of points that ought to be on the line that corresponds to that function. For example, what if $y = 0$? What value should x have? If $y = 0$, then a little arithmetic will reveal that x should equal -10. Thus, the pair $(0, -10)$ is a point on the line defined by the function. What if $x = 0$? What value will y have? In that case, the second

FIGURE 2.1
Graphs of the linear relationships $y = 5 + x/2$ (with a positive slope) and $y = 5 - x/2$ (with a negative slope).



term in the right-hand side of the equation disappears, so that $y = 5$. Thus, the pair of values (5, 0) is a point on the line defined by the function.

The graph of $y = 5 + x/2$ reveals some things about the relationship between y and x that we otherwise might not so easily discover. For example, notice that the line representing the equation slopes upward, or from southwest to northeast. The *positive slope*, as it is called, reveals that the relationship between x and y is a *direct* one. Thus, as x increases, so does y. And as x decreases, y decreases. Put more generally, when the independent and dependent variables move in the same direction, the slope of the graph of their relationship will be positive.

The graph also reveals the strength of this direct relationship by showing whether small changes in x lead to small or large changes in y. Notice that if x increases by 2 units, y increases by 1 unit. Another way of putting this is to say that in order to get a 10-unit increase in y, there must be a 20-unit increase in x.[3]

The opposite of a direct relationship is an *inverse* relationship. In that sort of relationship, the dependent and independent variables move in opposite directions. Thus, if x and y are inversely related, an *increase* in x (the independent variable) will lead to a *decrease* in y. Also, a *decrease* in x will lead to an *increase* in y. An example of an inverse relationship between an independent and a dependent variable is $y = 5 - x/2$. The graph of this line is also shown in Figure 2.1. Note that the line is downward-sloping; that is, the line runs from northwest to southeast.

> **QUESTION 2.1:** Suppose that the equation were $y = 5 + x$. Show in a graph like the one in Figure 2.1 what the graph of that equation would look like. Is the relationship between x and y direct or inverse? Is the slope of the new equation greater or less than the slope shown in Figure 2.1?
>
> Now suppose that the equation were $y = 5 - x$. Show in a graph like the one in Figure 2.1 what the graph of that equation would look like. Is the relationship between x and y direct or inverse? Is the slope of the new equation positive or negative? Would the slope of the equation $y = 5 - x/2$ be steeper or shallower than that of the one in $y = 5 - x$?

The graph of $y = 5 + x/2$ in Figure 2.1 also reveals that the relationship between the variables is *linear*. This means that when we graph the values of the independent and dependent variables, the resulting relationship is a straight line. One of the implications of linearity is that changes in the independent variable cause a constant rate of change in the dependent variable. In terms of Figure 2.1, if we would like to know the effect on y of doubling the amount of x, it doesn't matter whether we investigate that effect when x equals 2 or 3147. The effect on y of doubling the value of x is proportionally the same, regardless of the value of x.

The alternative to a linear relationship is, of course, a nonlinear relationship. In general, nonlinear relationships are trickier to deal with than are linear relationships.

---

[3] The slope of the equation we have been dealing with in Figure 2.1 is $\frac{1}{2}$, which is the coefficient of x in the equation. In fact, in any linear relationship the coefficient of the independent variable gives the slope of the equation.

FIGURE 2.2
The graph of a nonlinear relationship, given by the equation $y = x^2$.

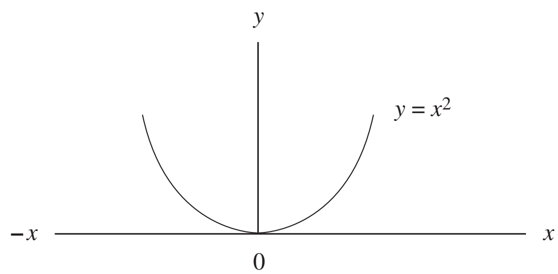
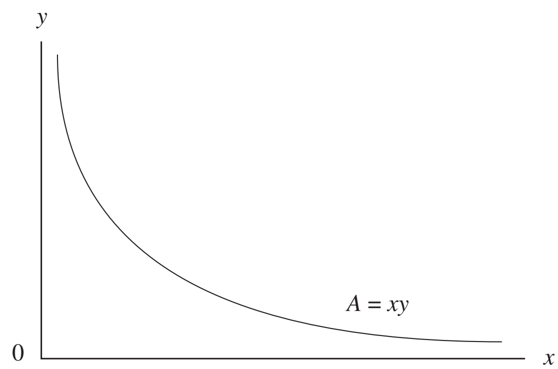


FIGURE 2.3
The graph of a nonlinear relationship, $A = xy$.



They frequently, although not always, are characterized by the independent variable being raised to a power by an exponent. Examples are $y = x^2$ and $y = 5/x^{\frac{1}{2}}$. Figure 2.2 shows a graph of $y = x^2$. Another common nonlinear relationship in economics is given by the example $A = xy$, where A is a constant. A graph of that function is given in Figure 2.3.

IV. The Theory of Consumer Choice and Demand

The economist's general theory of how people make choices is referred to as the theory of rational choice. In this section we show how that theory explains the consumer's choice of what goods and services to purchase and in what amounts.

A. Consumer Preference Orderings

The construction of the economic model of consumer choice begins with an account of the preferences of consumers. Consumers are assumed to know the things they like and dislike and to be able to rank the available alternative combinations of goods and services according to their ability to satisfy the consumer's preferences. This involves no more than ranking the alternatives as better than, worse than, or equally as good as one another. Indeed, some economists believe that the conditions they impose on the ordering or ranking of consumer preferences constitute what an economist means by the term *rational*. What are those conditions? They are that a consumer's preference ordering or ranking be *complete*, *transitive*, and *reflexive*. For an ordering to be *complete* simply means that the consumer be able to tell us how she ranks all the possible combinations of goods and services. Suppose that A represents a bundle of certain goods and services and B represents another bundle of the same goods and services but in different amounts. Completeness requires that the consumer be able to tell us that she prefers A to B, or that she prefers B to A, or that A and B are equally good (that is, that the consumer is indifferent between having A and having B). The consumer is *not* allowed to say, "I can't compare them."

Reflexivity is an arcane condition on consumer preferences. It means that any bundle of goods, A, is at least as good as itself. That condition is so trivially true that it is difficult to give a justification for its inclusion.

Transitivity means that the preference ordering obeys the following condition: If bundle A is preferred to bundle B and bundle B is preferred to bundle C, then it must be the case that A is preferred to C. This also applies to indifference: If the consumer is indifferent between A and B and between B and C, then she is also indifferent between A and C. Transitivity precludes the circularity of individual preferences. That is, transitivity means that it is impossible for A to be preferred to B, B to be preferred to C, and C to be preferred to A. Most of us would probably feel that someone who had circular preferences was extremely young or childish or crazy.

> **QUESTION 2.2:** Suppose that you have asked James whether he would like a hamburger or a hot dog for lunch, and he said that he wanted a hot dog. Five hours later you ask him what he would like for dinner, a hamburger or a hot dog. James answers, "A hamburger." Do James's preferences for hot dogs versus hamburgers obey the conditions above? Why or why not?

It is important to remember that the preferences of the consumer are *subjective*. Different people have different tastes, and these will be reflected in the fact that they may have very different preference orderings over the same goods and services. Economists leave to other disciplines, such as psychology and sociology, the study of the source of these preferences. We take consumer tastes or preferences as given, or, as economists say, as *exogenous*, which means that they are determined outside the economic system.[4]

An important consequence of the subjectivity of individual preferences is that economists have no accepted method for comparing the strength of people's preferences. Suppose that Stan tells us that he prefers bundle A to bundle B, and Jill tells us that she feels the same way: She also prefers A to B. Is there any way to tell who would prefer having A more? In the abstract, the answer is, "No, there is not." All we have from each consumer is the *order* of preference, not the *strength* of those preferences. Indeed, there is no metric by which to measure the strength of preferences, although economists sometimes jokingly refer to the "utils" of satisfaction that a consumer is enjoying. The inability to make *interpersonal comparisons of well-being* has some

[4] Many people new to the study of microeconomics will find this assumption of the exogeneity of preferences to be highly unrealistic. And there is some controversy about this assumption even within economics, some economists contending that preferences are endogenous—that is, determined within the economic system by such things as advertising. We cannot elaborate on this controversy here but are well aware of it.

important implications for the design and implementation of public policy, as we shall see in the section on welfare economics.

B. Utility Functions and Indifference Curves

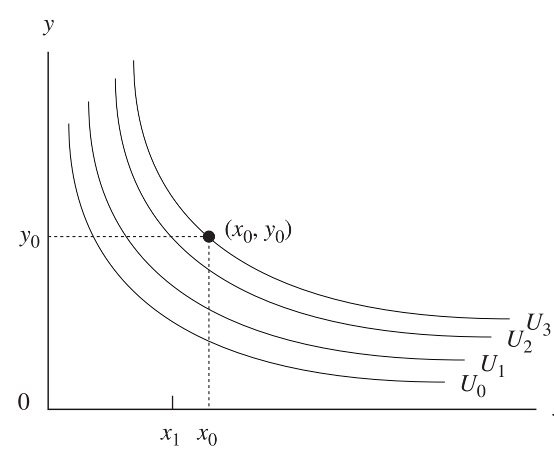
Once a consumer describes what his or her preference ordering is, we may derive a *utility function* for that consumer. The utility function identifies higher preferences with larger numbers. Suppose that there are only two commodities or services, x and y, available to a given consumer. If we let u stand for the consumer's utility, then the function $u = u(x, y)$ describes the utility that the consumer gets from different combinations of x and y.

A very helpful way of visualizing the consumer's utility function is by means of a graph called an *indifference map*. An example is shown in Figure 2.4. There we have drawn several *indifference curves*. Each curve represents all the combinations of x and y that give the consumer the same amount of utility or well-being. Alternatively, we might say that the consumer's tastes are such that he is indifferent among all the combinations of x and y that lie along a given curve—hence, the name *indifference curve*. Thus, all those combinations of x and y lying along the indifference curve marked U_0 give the consumer the same utility. Those combinations lying on the higher indifference curve marked U_1 give this consumer similar utility, but this level of utility is higher than that of all those combinations of x and y lying along indifference curve U_0.

> **QUESTION 2.3:** Begin at point (x_0, y_0). Now decrease x from x_0 to x_1. How much must y increase to offset the decrease in x and keep the consumer indifferent?

The problem of consumer choice arises from the collision of the consumer's preferences with obstacles to his or her satisfaction. The obstacles are the constraints that force decision makers to choose among alternatives. There are many constraints, including time, energy, knowledge, and one's culture, but foremost among these is

FIGURE 2.4
The consumer's indifference map.



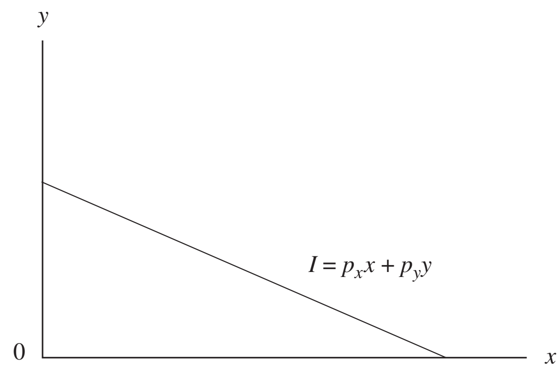


FIGURE 2.5
The consumer's income constraint or budget line.

limited income. We can represent the consumer's income constraint or budget line by the line in Figure 2.5. The area below the line and the line itself represent all the combinations of x and y that are affordable, given the consumer's income, I.[5] Presumably, the consumer intends to spend all of her income on purchases of these two goods and services, so that the combinations upon which we shall focus are those that are on the budget line itself.

QUESTION 2.4: In a figure like the one in Figure 2.5 and beginning with a budget line like the one in Figure 2.5, show how you would draw the new income constraint to reflect the following changes?

1. An increase in the consumer's income, prices held constant.
2. A decrease in the consumer's income, prices held constant.
3. A decrease in the price of x, income and the price of y held constant.
4. An increase in the price of y, income and the price of x held constant.

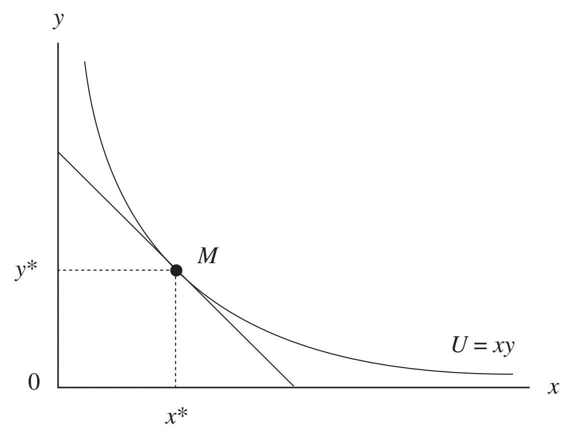
C. The Consumer's Optimum

We may now combine the information about the consumer's tastes given by the indifference map and the information about the income constraint given by the budget line in order to show what combination of x and y maximizes the consumer's utility, subject to the constraint imposed by her income. See Figure 2.6. There the consumer's optimum bundle is shown as point M, which contains x^* and y^*. Of all the feasible combinations of x and y, that combination gives this consumer the greatest utility.[6]

[5] The equation for the budget line is $I = p_x x + p_y y$, where p_x is the price per unit of x and p_y is the price per unit of y. As an exercise, you might try to rearrange this equation, with y as the dependent variable, in order to show that the slope of the line is negative. When you do so, you will find that the coefficient of the x-term is equal to $-p_x/p_y$. Economists refer to this ratio as *relative price*.

[6] Because we have assumed that the normal indifference curves are convex to the origin, there is a *unique* bundle of x and y that maximizes the consumer's utility. For other shapes of the indifference curves it is possible that there is more than one bundle that maximizes utility.

FIGURE 2.6
The consumer's optimum.



D. A Generalization: The Economic Optimum as Marginal Cost = Marginal Benefit

Because of the central importance of constrained maximization in microeconomic theory, let us take a moment to examine a more general way of characterizing such a maximum:

> A constrained maximum, or any other economic optimum, can be described as a point where marginal cost equals marginal benefit.

Let's see how this rule characterizes maximizing decisions.[7] Begin by assuming that the decision maker chooses some initial level of whatever it is he is interested in maximizing. He then attempts to determine whether that initial level is his maximum; is that level as good as he can do, given his constraints? He can answer the question by making very small, what an economist calls *marginal,* changes away from that initial level. Suppose that the decision maker proposes to *increase* slightly above his initial level whatever it is he is doing. There will be a cost associated with this small increase called *marginal cost*. But there will also be a benefit of having or doing more of whatever it is that he is attempting to maximize. The benefit of this small increase is called *marginal benefit*. The decision maker will perceive himself as doing better at this new level, by comparison to his initial level, so long as the *marginal benefit* of the small increase is greater than the *marginal cost* of the change. He will continue to make these small, or marginal, adjustments so long as the marginal benefit exceeds the marginal cost, and he will stop making changes when the marginal cost of the last change made equals (or is greater than) the marginal benefit. That level is the decision maker's maximum.

QUESTION 2.5: Suppose that, instead of increasing his level above the initial choice, the decision maker first tries decreasing the amount of whatever it is he is attempting to maximize. Explain how the comparison of marginal cost

[7] This rule could describe equally well an economic optimum where the goal of the decision maker is to *minimize* something. In that case, the optimum would still be the point at which $MC = MB$, but the demonstration of the stylized decision making that got one to that point would be different from that given in the text.

and marginal benefit for these decreases is made and leads the decision maker to the optimum. (Assume that the initial level is greater than what will ultimately prove to be the optimum.)

We can characterize the consumer's income-constrained maximum, *M* in Figure 2.6, in terms of the equality of marginal cost and benefit. Small changes in either direction along the budget line, *I*, represent a situation in which the consumer spends a dollar less on one good and a dollar more on the other. To illustrate, assume the consumer decides to spend a dollar less on *y* and a dollar more on *x*. Purchasing a dollar less of *y* causes a loss in utility that we may call the marginal cost of the budget reallocation. But the dollar previously spent on *y* can now be spent on *x*. More units of *x* mean greater utility, so that we may call this increase the marginal benefit of the budget reallocation.

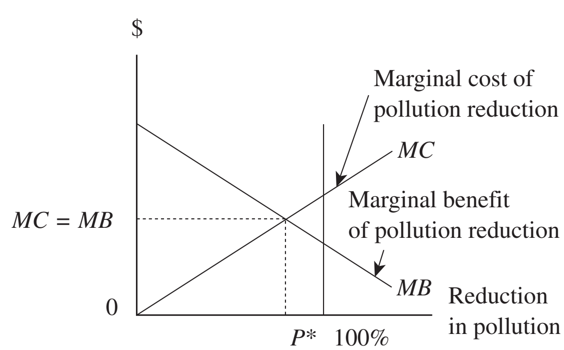
Should the consumer spend a dollar less on good *y* and a dollar more on *x*? Only if the marginal cost (the decrease in utility from one dollar less of *y*) is less than the marginal benefit (the increase in utility from having one dollar more of *x*). The rational consumer will continue to reallocate dollars away from the purchase of *y* and toward the purchase of *x* until the marginal benefit of the last change made equals the marginal cost. This occurs at the point *M* in Figure 2.6.

Figure 2.7 applies constrained maximization to reduce the amount of pollution. Along the vertical axis are dollar amounts. Along the horizontal axis are units of pollution reduction. At the origin there is no effort to reduce pollution. At the vertical line labeled "100%," pollution has been completely eliminated.

The curve labeled *MB* shows the marginal benefit to society of reducing pollution. We assume that this has been correctly measured to take into account health, scenic, and all other benefits that accrue to members of society from reducing pollution at various levels. This line starts off high and then declines. This downward slope captures the fact that the very first efforts at pollution reduction confer large benefits on society. The next effort at reducing pollution also confers a social benefit, but not quite as great as the initial efforts. Finally, as we approach the vertical line labeled "100%" and all vestiges of pollution are being eliminated, the benefit to society of achieving those last steps is positive, but not nearly as great as the benefit of the early stages of pollution reduction.

FIGURE 2.7

The socially optimal amount of pollution-reduction effort.



The curve labeled *MC* represents the "social" as opposed to "private" marginal cost of achieving given levels of pollution reduction. The individuals and firms that pollute must incur costs to reduce pollution: They may have to adopt cleaner and safer production processes that are also more expensive; they may have to install monitoring devices that check the levels of pollution they generate; and they may have to defend themselves in court when they are accused of violating the pollution-reduction guidelines. We have drawn the *MC* curve to be upward-sloping to indicate that the marginal costs of achieving any given level of pollution-reduction increase. This means that the cost of reducing the very worst pollution may not be very high, but that successive levels of reduction will be ever more expensive.

Given declining marginal benefit and rising marginal cost, the question then arises, "What is the optimal amount of pollution-reduction effort for society?" An examination of Figure 2.7 shows that P^* is the socially optimal amount of pollution-reduction effort. Any more effort will cost more than it is worth. Any less would cause a reduction in benefits that would be greater than the savings in costs.

Note that, according to this particular graph, it would not be optimal for society to try to eliminate pollution entirely. Here it is socially optimal to tolerate some pollution. Specifically, when pollution reduction equals P^*, the remaining pollution equals $100\% - P^*$, which is the "optimal amount of pollution." Few goods are free. Much of the wisdom of economics comes from the recognition of this fact and of the derivation of techniques for computing the costs and benefits.

> **QUESTION 2.6:** Suppose that we were to characterize society's decision making with regard to pollution-reduction efforts as an attempt to maximize the *net benefit* of pollution-reduction efforts. Let us define *net benefit* as the difference between marginal benefit and marginal cost. What level of pollution-reduction effort corresponds to this goal?

> **QUESTION 2.7:** Using a graph like Figure 2.7, show the effect on the determination of the socially optimal amount of pollution-reduction effort of the following:
>
> 1. Some technological change that lowers the marginal cost of achieving every level of pollution reduction.
> 2. A discovery that there are greater health risks associated with every given level of pollution than were previously thought to be the case.

If you understand that for economists, *the optimum for nearly all decisions occurs at the point at which marginal benefit equals marginal cost*, then you have gone a long way toward mastering the microeconomic tools necessary to answer most questions that we will raise in this book.

E. Individual Demand

We may use the model of consumer choice of the previous sections to derive a relationship between the price of a good and the amount of that good in a consumer's optimum bundle. The *demand curve* represents this relationship.

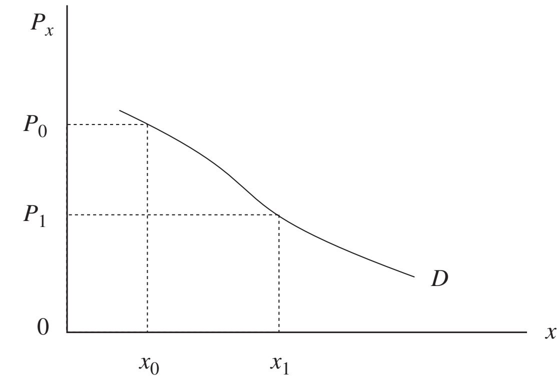


FIGURE 2.8
An individual's demand curve, showing the inverse relationship between price and quantity demanded.

Starting from point M in Figure 2.6, note that when the price of x is that given by the budget line, the optimal amount of x to consume is x^*. But what amount of x will this consumer want to purchase so as to maximize utility when the price of x is lower than that given by the budget line in Figure 2.6? We can answer that question by holding P_y and I constant, letting P_x fall, and writing down the amount of x in the succeeding optimal bundles. Not surprisingly, the result of this exercise will be that the price of x and the amount of x in the optimum bundles are inversely related. That is, when the price of x goes up, P_y and I held constant (or *ceteris paribus*, "all other things equal," as economists say), the amount of x that the consumer will purchase goes down, and vice versa. This result is the famous *law of demand*.

We may graph this relationship between P_x and the quantity of x demanded to get the individual demand curve, D, shown in Figure 2.8. The demand curve we have drawn in Figure 2.8 could have had a different slope than that shown; it might have been either flatter or steeper. The steepness of the demand curve is related to an important concept called the *price elasticity of demand*, or simply *elasticity of demand*.[8]

This is an extremely useful concept: It measures how responsive consumer demand is to changes in price. And there are some standard attributes of goods that influence how responsive demand is likely to be. For instance, if two goods are similar in their use, then an increase in the price of the first good with no change in the price of the second good causes consumers to buy significantly less of the first good. Generalizing, the most important determinant of the price elasticity of demand for a

[8] The measure is frequently denoted by the letter e, and the ranges of elasticity are called *inelastic* ($e < 1$), *elastic* ($e > 1$), and *unitary elastic* ($e = 1$). By convention, e, the price elasticity of demand, is a positive (or absolute) number, even though the calculation we suggested will lead to a negative number. For an inelastically demanded good, the percentage change in price exceeds the percentage change in quantity demanded. Thus, a good that has $e = 0.5$ is one for which a 50 percent decline in price will cause a 25 percent increase in the quantity demanded, or for which a 15 percent increase in price will cause a 7.5 percent decline in quantity demanded. For an elastically demanded good, the percentage change in price is less than the percentage change in quantity demanded. As a result, a good that has $e = 1.5$ is one for which a 50 percent decline in price will cause a 75 percent increase in quantity demanded, or for which a 20 percent increase in price will cause a 30 percent decline in quantity demanded.

good is the availability of substitutes. The more substitutes for the good, the greater the elasticity of demand; the fewer the substitutes, the lower the elasticity. Substitution is easier for narrowly defined goods and harder for broad categories. If the price of cucumbers goes up, switching to peas or carrots is easy; if the price of vegetables goes up, switching to meat is possible; but if the price of food goes up, eating less is hard to do. So, we expect that demand is more elastic for cucumbers than vegetables and more elastic for vegetables than food. Also, demand is more elastic in the long run than the short run. To illustrate, if electricity prices rise relative to natural gas, consumers will increasingly switch to burning gas as they gradually replace furnaces and appliances. Economists often measure and remeasure the price elasticities of demand for numerous goods and services to predict responses to price changes.

V. The Theory of Supply

We now turn to a review of the other side of the market: the supply side. The key institution in supplying goods and services for sale to consumers is the business firm. In this section we shall see what goal the firm seeks and how it decides what to supply. In the following section, we merge our models of supply and demand to see how the independent maximizing activities of consumers and firms achieve a market equilibrium.

A. The Profit-Maximizing Firm

The firm is the institution in which output (products and services) is fabricated from inputs (capital, labor, land, and so on). Just as we assume that consumers rationally maximize utility subject to their income constraint, we assume that firms *maximize profits subject to the constraints imposed on them by consumer demand and the technology of production*.

In microeconomics, *profits* are defined as the difference between *total revenue* and the *total costs* of production. Total revenue for the firm equals the number of units of output sold multiplied by the price of each unit. Total costs equal the costs of each of the inputs times the number of units of input used, summed over all inputs. The profit-maximizing firm produces that amount of output that leads to the greatest positive difference between the firm's revenue and its costs. Microeconomic theory demonstrates that the firm will maximize its profits if it produces that *amount of output whose marginal cost equals its marginal revenue*. (In fact, this is simply an application of the general rule we discussed in section IV.D earlier: To achieve an optimum, equate marginal cost and marginal benefit.)

These considerations suggest that when marginal revenue exceeds marginal cost, the firm should expand production, and that when marginal cost exceeds marginal revenue, it should reduce production. It follows that profits will be maximized for that output for which marginal cost and marginal revenue are equal. Note the economy of this rule: To maximize profits, the firm need not concern itself with its total costs or total revenues; instead, it can simply experiment on production unit by unit in order to discover the output level that maximizes its profits.

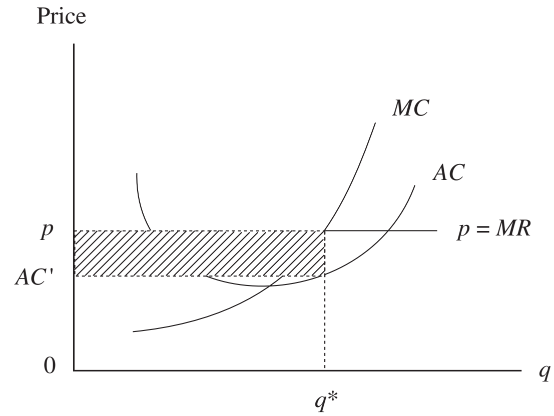


FIGURE 2.9
The profit-maximizing output for a firm.

In Figure 2.9 the profit-maximizing output of the firm is shown at the point at which the marginal cost curve, labeled *MC*, and marginal revenue curve of the firm are equal. The profit-maximizing output level is denoted q^*. Total profits at this level of production, denoted by the shaded area in the figure, equal the difference between the total revenues of the firm (p times q^*) and the total costs of the firm (the average cost of producing q^* times q^*).

There are several things you should note about the curves in the graph. We have drawn the marginal revenue curve as horizontal and equal to the prevailing price. This implies that the firm can sell as much as it likes at that prevailing price. Doubling its sales will have no effect on the market price of the good or service. This sort of behavior is referred to as *price-taking* behavior. It characterizes industries in which there are so many firms, most of them small, that the actions of no single firm can affect the market price of the good or service. An example might be farming. There are so many suppliers of wheat that the decision of one farmer to double or triple output or cut it in half will have no impact on its market price. (Of course, if all farms decide to double output, there will be a substantial impact on market price.) Such an industry is said to be "perfectly competitive."

B. The Short Run and the Long Run

In microeconomics the firm is said to operate in two different time frames: the short run and the long run. These time periods do not correspond to calendar time. Instead they are defined in terms of the firm's inputs. In the short run at least one input is fixed (all others being variable), and the usual factor of production that is fixed is capital (the firm's buildings, machines, and other durable inputs). Because capital is fixed in the short run, all the costs associated with capital are called *fixed costs*. In the short run the firm can, in essence, ignore those costs: They will be incurred regardless of whether the firm produces nothing at all or 10 million units of output. (The only costs that change in the short run are "variable costs," which rise or fall depending on how much output the firm produces.) The long run is distinguished by the fact that all factors of production become variable. There are no longer any fixed costs. Established firms may expand their productive capacity or leave the industry entirely, and new firms may enter the business.

Another important distinction between the long and the short run has to do with the equilibrium level of profits for each firm. At any point in time there is an average rate of return earned by capital in the economy as a whole. When profits being earned in a particular industry exceed the average profit rate for comparable investments, firms will enter the industry, assuming there are no barriers to entry. As entry occurs, the total industry output increases, and the price of the industry output goes down, causing each firm's revenue to decrease. Also, the increased competition for the factors of production causes input prices to rise, pushing up each firm's costs. The combination of these two forces causes each firm's profits to decline. Entry ceases when profits fall to the average rate.

Economists have a special way of describing these facts. The average return on capital is treated as part of the costs that are subtracted from revenues to get "economic profits." Thus, when the rate of return on invested capital in this industry equals the average for the economy as a whole, it is said that "economic profits are zero."[9]

This leads to the conclusion that economic profits are zero in an industry that is in long-run equilibrium. Because this condition can occur only at the minimum point of the firm's average cost curve, where the average costs of production are as low as they can possibly be, inputs will be most efficiently used in long-run equilibrium. Thus, the condition of zero economic profits, far from being a nightmare, is really a desirable state.

VI. Market Equilibrium

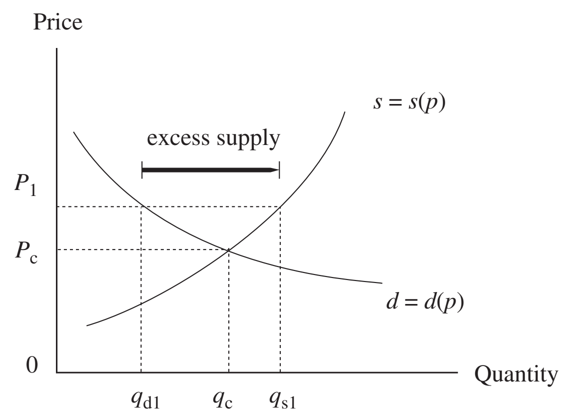
Having described the behavior of utility-maximizing consumers and profit-maximizing producers, our next task is to bring them together to explain how they interact. We shall first demonstrate how a unique price and quantity are determined by the interaction of supply and demand in a perfectly competitive market and then show what happens to price and quantity when the market structure changes to one of monopoly. We conclude this section with an example of equilibrium analysis of an important public policy issue.

A. Equilibrium in a Perfectly Competitive Industry

An industry in which there are so many firms that no one of them can influence the market price by its individual decisions and in which there are so many consumers that the individual utility-maximizing decisions of no one consumer can affect the market price is called a *perfectly competitive industry*. For such an industry the aggregate demand for and the aggregate supply of output can be represented by the downward-sloping demand curve, $d = d(p)$, and the upward-sloping supply curve, $s = s(p)$, shown

[9] When profits in a given industry are less than the average in the economy as a whole, economic profits are said to be negative. When that is the case, firms exit this industry for other industries where the profits are at least equal to the average for the economy. As an exercise, see if you can demonstrate the process by which profits go to zero when negative economic profits in an industry cause exit to take place.

FIGURE 2.10
Market equilibrium in a perfectly competitive market.



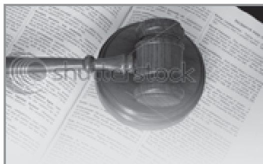
in Figure 2.10. The *market-clearing* or *equilibrium* price and quantity occur at the point of intersection of the aggregate supply and demand curves. At that combination of price and quantity, the decisions of consumers and suppliers are consistent.

One way to see why the combination P_c, q_c in Figure 2.10 is an equilibrium is to see what would happen if a different price-quantity combination were obtained. Suppose that the initial market price was P_1. At that price, producers would maximize their profits by supplying q_{s1} of output, and utility-maximizing consumers would be prepared to purchase q_{d1} units of output. These supply and demand decisions are inconsistent: At P_1, the amount that suppliers would like to sell exceeds the amount that consumers would like to buy. How will the market deal with this excess supply? Clearly, the market price must fall. As the price falls, consumers will demand more and producers will supply less, so the gap between supply and demand will diminish. Eventually the price may reach P_c. And at that price, as we have seen, the amount that suppliers wish to sell and the amount that consumers wish to purchase are equal.

B. Equilibrium in a Monopolistic Market

Monopoly is at the other extreme of market structure. In a monopoly there is only one supplier; so, that firm and the industry are identical. A monopoly can arise and persist only where there are barriers to entry that make it impossible for competing firms to appear. In general, such barriers can arise from two sources: first, from statutory and other legal restrictions on entry; and second, from technological conditions of production known as *economies of scale*. An example of a statutory restriction on entry was the Civil Aeronautics Board's refusal from the 1930s until the mid-1970s to permit entry of new airlines into the market for passenger traffic on such major routes as Los Angeles–New York and Chicago–Miami.

The second barrier to entry is technological. *Economies of scale* are a condition of production in which the greater the level of output, the lower the average cost of production. Where such conditions exist, one firm can produce any level of output at less cost than multiple firms. A monopolist that owes its existence to economies of scale is sometimes called a *natural monopoly*. Public utilities, such as local water, telecommunications, cable, and power companies, are often natural monopolies. The technological advantages of a natural monopoly would be partially lost if the single firm is allowed



Opportunity Cost and Comparative Advantage

We have been implicitly using one of the most fundamental concepts in microeconomics: *opportunity cost*. This term refers to the economic cost of an alternative that has been foregone. When you decided to attend a college, graduate school, or law school, you gave up certain other valuable alternatives, such as taking a job, training for the Olympics, or traveling around the world on a tramp steamer. In reckoning the cost of going to college, graduate school, or law school, the true economic cost was that of the next best alternative. This point is true of the decisions of all economic actors: When maximizing utility, the consumer must consider the opportunities given up by choosing one bundle of consumer goods rather than another; when maximizing profits, the firm must consider the opportunities foregone by committing its resources to the production of widgets rather than to something else.

In general, the economic notion of opportunity cost is more expansive than the more common notion of accounting cost. An example will make this point.[10] Suppose that a rich relative gives you a car whose market value is $15,000. She says that if you sell the car, you may keep the proceeds, but that if you use the car yourself, she'll pay for the gas, oil, maintenance, repairs, and insurance. In short she says, "The use of the car is FREE!" But is it? Suppose that the $15,000 for which the car could be sold would earn 12 percent interest per year in a savings account, giving $1800 per year in interest income. If you use the car for 1 year, its resale value will fall to $11,000—a cost to you of $4000. Therefore, the opportunity cost to you of using the car for 1 year is $4000 plus the foregone interest of $1800—a total of $5800. This is far from being free. The accounting cost of using the car is zero, but the opportunity cost is positive.

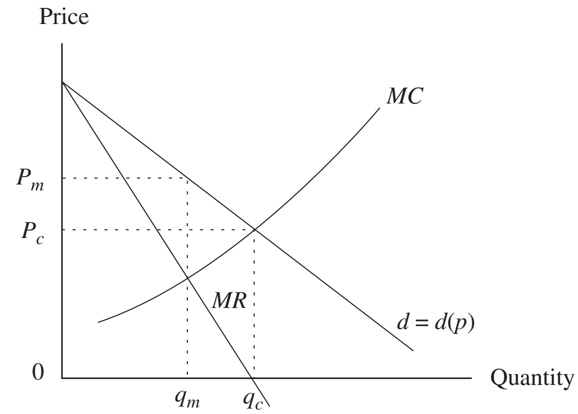
Comparative advantage is another useful economic concept related to the notion of opportunity cost. The law of comparative advantage asserts that people should engage in those pursuits where their opportunity costs are lower than others. For example, someone who is 7 feet tall has a comparative advantage in pursuing a career in professional basketball. But what about someone whose skills are such that she can do many things well? Suppose, for example, that a skilled attorney is also an extremely skilled typist. Should she do her own typing or hire someone else to do it while she specializes in the practice of law? The notion of comparative advantage argues for specialization: The attorney can make so much more money by specializing in the practice of law than by trying to do both jobs that she could easily afford to hire someone else who is less efficient at typing to do her typing for her.

to restrict its output and to charge a monopoly price. For that reason, natural monopolies are typically regulated by the government.

The monopolist, like the competitive firm, maximizes profit by producing that output for which marginal cost equals marginal revenue. Marginal cost of the monopolist, as for the competitive firm, is the cost of producing one more unit of output. This cost curve is represented in Figure 2.11 by the curve labeled *MC*. But marginal revenue for the monopolist is different from what it was for the competitive firm. Recall that marginal revenue describes the change in a firm's total revenues for a small, or marginal,

[10] The example is taken from Roy Ruffin & Paul Gregory, Principles of Microeconomics 156 (2d ed. 1986).

FIGURE 2.11
Profit-maximizing output and price for a monopolist.



change in the number of units of output sold. For the competitive firm marginal revenue is equal to the price of output. Because the competitive firm can sell as much as it likes at the prevailing price, each additional unit of output sold adds exactly the sale price to the firm's total revenues. But for the monopolist, marginal revenue declines as the number of units sold increases. This is indicated in Figure 2.11 by the downward-sloping curve labeled *MR*. Notice that the *MR* curve lies below the demand curve. This indicates that the marginal revenue from any unit sold by a monopolist is always less than the price. *MR* is positive but declining for units of output between 0 and q_c; thus, the sale of each of those units increases the firm's total revenues but at a decreasing rate. The q_0 unit actually adds nothing to the firm's total revenues $(MR = 0)$, and for each unit of output beyond q_c, *MR* is less than zero, which means that each of those units actually reduces the monopolist's total revenues.

The reason for this complex relationship between marginal revenue and units sold by the monopolist is the downward-sloping demand curve. The downward-sloping demand curve implies that the monopolist must lower the price to sell more units; but in order to sell an additional unit of output he or she must lower the price not just on the last or marginal unit but on all the units sold.[11] From this fact it can be shown, using calculus, that the addition to total revenues from an additional unit of output sold will always be less than the price charged for that unit. Thus, because *MR* is always less than the price for all units of output and because price declines along the demand curve, the *MR* curve must also be downward sloping and lie below the demand curve.

The monopolist maximizes his profit by choosing that output level for which marginal revenue and marginal cost are equal. This output level, q_m, is shown in Figure 2.11. The demand curve indicates that consumers are willing to pay P_m for that amount of output. Notice that if this industry were competitive instead of monopolized, the profit-maximizing actions of the firms would have resulted in an equilibrium price and quantity at the intersection of the aggregate supply curve, *S*, and the industry demand curve, *D*. The competitive price, P_c, is lower than the monopolistic price, and the

[11] This assumes that the monopolist cannot price-discriminate (that is, charge different prices to different consumers for the same product).

quantity of output produced and consumed under competition, q_c, is greater than under monopoly.

Economists distinguish additional market structures that are intermediate between the extremes of perfect competition and monopoly. The most important among these are *oligopoly* and *imperfect competition*. An oligopolistic market is one containing a few firms that recognize that their individual profit-maximizing decisions are interdependent. That means that what is optimal for firm *A* depends not only on its marginal costs and the demand for its output but also on what firms *B*, *C*, and *D* have decided to produce and the prices they are charging. The economic analysis of this interdependence requires a knowledge of game theory, which we discuss below.

An imperfectly competitive market is one that shares most of the characteristics of a perfectly competitive market—for example, free entry and exit of firms and the presence of many firms—but has one important monopolistic element: Firms produce differentiable output rather than the homogeneous output produced by perfectly competitive firms. Thus, imperfectly competitive firms distinguish their output by brand names, colors, sizes, quality, durability, and so on.

C. An Example of Equilibrium Analysis

It is useful to have an example applying this theory to a real problem. Let us imagine a market for rental housing like the one shown in Figure 2.12. The demand for rental housing is given by the curve *D*, and the supply of rental housing is given by the upward-sloping supply curve *S*. Assuming that the rental housing market is competitive, then the independent actions of consumers and of profit-maximizing housing owners will lead to a rental rate of r_1 being charged and of h_1 units of rental housing being supplied and demanded. Note that this is an equilibrium in the sense we discussed above: The decisions of those demanding the product and of those supplying it are consistent at the price r_1. Unless something causes the demand curve or the supply curve to shift, this price and output combination will remain in force.

But now suppose that the city government feels that r_1 is too high and passes an ordinance that specifies a maximum rental rate for housing of r_m, considerably below

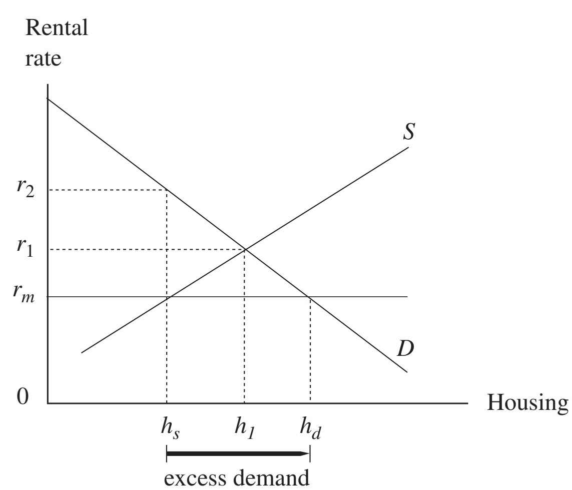


FIGURE 2.12

The consequences of a rent-control ordinance that prescribes rents below the market-clearing rental rate.

the equilibrium market rate. The hope of the government is that at least the same amount of housing will be consumed by renters but at a lower rental rate. A look at Figure 2.12, however, leads one to doubt that result. At r_m, consumers demand h_d units of rental housing, an increase over the quantity demanded at the higher rate, r_1. But at this lower rate suppliers are only prepared to supply h_s units of rental housing. Apparently it does not pay them to devote as much of their housing units to renters at that lower rate; perhaps if r_m is all one can get from renting housing units, suppliers prefer to switch some of their units to other uses, such as occupancy by the owner's family or their sale as condominiums. The result of the rate ceiling imposed by the government is a shortage of, or excess demand for, rental units equal to $(h_d - h_s)$.

If the rate ceiling is strictly enforced, the shortage will persist. Some non-price methods of determining who gets the h_s units of rental housing must be found, such as queuing. Eventually, the shortage may be eased if either the demand curve shifts inward or the supply curve shifts outward. It is also possible that landlords will let their property deteriorate by withholding routine maintenance and repairs, so that the quality of their property falls to such an extent that r_m provides a competitive rate of return to them.

If, however, the rate ceiling is *not* strictly enforced, then consumers and suppliers will find a way to erase the shortage. For example, renters could offer free services or secret payments (sometimes called *side payments*) to landlords in order to get the effective rental rate above r_m and induce the landlord to rent to them rather than to those willing to pay only r_m. Those services and side payments could amount to $(r_2 - r_m)$ per housing unit.

VII. Game Theory

The law frequently confronts situations in which there are few decision makers and in which the optimal action for one person to take depends on what another actor chooses. These situations are like games in that people must decide upon a strategy. A strategy is a plan for acting that responds to the reactions of others. *Game theory* deals with any situation in which strategy is important. Game theory will, consequently, enhance our understanding of some legal rules and institutions. For those who would like to pursue this topic in more detail, there are now several excellent introductory books on game theory.[12]

To characterize a game, we must specify three things:

1. *players,*
2. *strategies* of each player, and
3. *payoffs* to each player for each strategy.

---

[12] For those who would like to pursue game theory in more detail, there are now several excellent introductory texts: ERIC RASMUSEN, GAMES AND INFORMATION: AN INTRODUCTION TO GAME THEORY (3d ed. 2001); DAVID KREPS, GAME THEORY AND ECONOMIC MODELING (1990); and AVINASH DIXIT & BARRY NALEBUFF, THINKING STRATEGICALLY: THE COMPETITIVE EDGE IN BUSINESS, POLITICS, AND EVERYDAY LIFE (1991). More advanced treatments may be found in ROGER MYERSON, GAME THEORY (1991) and DREW FUDENBERG & JEAN TIROLE, GAME THEORY (1991). With special reference to law, see DOUGLAS BAIRD, ROBERT GERTNER, & RANDAL PICKER, GAME THEORY AND THE LAW (1995).

Let's consider a famous example—the prisoner's dilemma. Two people, *Suspect 1* and *Suspect 2*, conspire to commit a crime. They are apprehended by the police outside the place where the crime was committed, taken to the police station, and placed in separate rooms so that they cannot communicate. The authorities question them individually and try to play one suspect against the other. The evidence against them is circumstantial—they were simply in the wrong place at the wrong time. If the prosecutor must go to trial with only this evidence, then the suspects will have to be charged with a minor offense and given a relatively light punishment—say, 1 year in prison. The prosecutor would very much prefer that one or both of the suspects confesses to the more serious crime that they are thought to have committed. Specifically, if either suspect confesses (and thereby implicates the other) and the other does not, the non-confessor will receive 7 years in prison, and as a reward for assisting the state, the confessor will only receive six months in jail. If both suspects can be induced to confess, each will spend 5 years in prison. What should each suspect do—confess or keep quiet?

The strategies available to the suspects can be shown in a *payoff matrix* like that in Figure 2.13. Each suspect has two strategies: confess or keep quiet. The payoffs to each player from following a given strategy are shown by the entries in the four cells of the box, with the payoff to *Suspect 2* given in the lower left-hand corner of each cell and the payoff to *Suspect 1* given in the upper right-hand corner of the cell.

Here is how to read the entries in the payoff matrix. If *Suspect 1* confesses and *Suspect 2* also confesses, each will receive 5 years in prison. If *Suspect 1* confesses and *Suspect 2* keeps quiet, *Suspect 1* will spend six months in jail, and *Suspect 2* will spend 7 years in prison. If *Suspect 1* keeps quiet and *Suspect 2* confesses, then *Suspect 2* will spend six months in jail, and *Suspect 1* will spend 7 years in prison. Finally, if both suspects keep quiet, each will spend 1 year in prison.

There is another way to look at *Suspect 1*'s options. The payoff matrix is sometimes referred to as the *strategic form* of the game. An alternative is the *extensive form*. This puts one player's options in the form of a decision tree, which is shown in Figure 2.14.

We now wish to explore what the optimal strategy—confess or keep quiet—is for each player, given the options in the payoff matrix and given some choice made by the other player. Let's consider how *Suspect 1* will select her optimal strategy. Remember

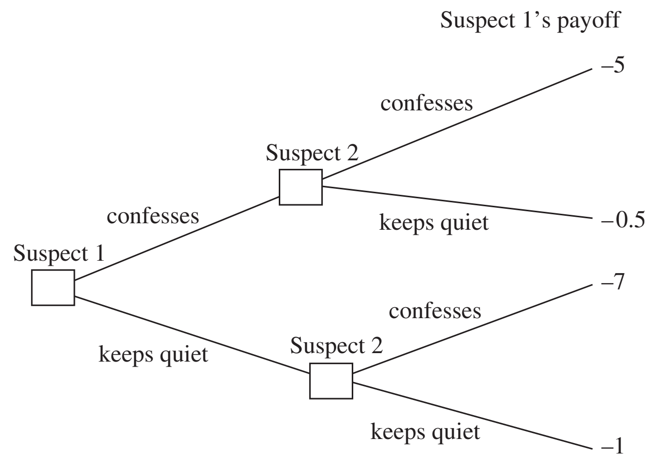
FIGURE 2.13

The strategic form of a game, also known as a payoff matrix.

		Suspect 1	
		Confess	Keep quiet
Suspect 2	Confess	–5 –5	–7 –0.5
	Keep quiet	–0.5 –7	–1 –1

FIGURE 2.14

The extensive form of the prisoner's dilemma.



that the players are being kept in separate rooms and cannot communicate with one another. (Because the game is symmetrical, this is exactly the same way in which *Suspect 2* will select his optimal strategy.)

First, what should *Suspect 1* do if *Suspect 2* confesses? If she keeps quiet when *Suspect 2* confesses, she will spend 7 years in prison. If she confesses when *Suspect 2* confesses, she will spend 5 years. So, if *Suspect 2* confesses, clearly the best thing for *Suspect 1* to do is to confess.

But what if *Suspect 2* adopts the alternative strategy of keeping quiet? What is the best thing for *Suspect 1* to do then? If *Suspect 2* keeps quiet and *Suspect 1* confesses, she will spend only half a year in prison. If she keeps quiet when *Suspect 2* keeps quiet, she will spend 1 year in prison. Again, the best thing for *Suspect 1* to do if the other suspect keeps quiet is to confess.

Thus, *Suspect 1* will always confess. Regardless of what the other player does, confessing will always mean less time in prison for her. In the jargon of game theory this means that confessing is a *dominant strategy*—the optimal move for a player to make is the same, regardless of what the other player does.

Because the other suspect will go through precisely the same calculations, he will also confess. Confessing is the dominant strategy for each player. The result is that the suspects are *both* going to confess, and, therefore, each will spend 5 years in prison.

The solution to this game, that both suspects confess, is an equilibrium: There is no reason for either player to change his or her strategy. There is a famous concept in game theory that characterizes this equilibrium—a *Nash equilibrium*. In such an equilibrium, no individual player can do any better by changing his or her behavior so long as the other players do not change theirs. (Notice that the competitive equilibrium that we discussed in previous sections is an example of a Nash equilibrium when there are many players in the game.)

The notion of a Nash equilibrium is fundamental in game theory, but it has shortcomings. For instance, some games have no Nash equilibrium. Some games have several Nash equilibria. And finally, there is not necessarily a correspondence between the

Nash equilibrium and Pareto efficiency, the criterion that economists use to evaluate many equilibria. To see why, return to the prisoner's dilemma above. We have seen that it is a Nash equilibrium for both suspects to confess. But you should note that this is *not* a Pareto-efficient solution to the game from the viewpoint of the accused. When both suspects confess, they will each spend 5 years in prison. It is possible for *both* players to be better off. That would happen if they would both keep quiet. Thus, cell 4 (where each receives a year in prison) is a Pareto-efficient outcome. Clearly, that solution is impossible because the suspects cannot make binding commitments not to confess.[13]

We may use the prisoner's dilemma to discuss another important fundamental concept of game theory—*repeated games*. Suppose that the prisoner's dilemma were to be played not just once but a number of times by the same players. Would that change our analysis of the game? If the same players play the same game according to the same rules repeatedly, then it is possible that cooperation can arise and that players have an incentive to establish a reputation—in this case, for trustworthiness.

An important thing to know about a repeated game is whether the game will be repeated a *fixed* number of times or an *indefinite* number. To see the difference, suppose that the prisoner's dilemma above is to be repeated exactly ten times. Each player's optimal strategy must now be considered across games, not just for one game at a time. Imagine *Suspect 2* thinking through, before the first game is played, what strategy he ought to follow for each game. He might imagine that he and his partner, if caught after each crime, will learn (or agree) to keep quiet rather than to confess. But then *Suspect 2* thinks forward to the final game, the tenth. Even if the players had learned (or agreed) to keep quiet through Game 9, things will be different in Game 10. Because this is the last time the game is to be played, *Suspect 1* has a strong incentive to confess. If she confesses on the last game and *Suspect 2* sticks to the agreement not to confess, he will spend 7 years in prison to her half year. Knowing that she has this incentive to cheat on an agreement not to confess in the last game, the best strategy for *Suspect 2* is also to confess in the final game. But now Game 9 becomes, in a sense, the final game. And in deciding on the optimal strategy for that game, exactly the same logic applies as it did for Game 10—both players will confess in Game 9, too. *Suspect 1* can work all this out, too, and she will realize that the best thing to do is to confess in Game 8, and so on. In the terminology of game theory, the game *unravels* so that confession takes place by each player every time the game is played, *if it is to be played a fixed number of times*.

Things may be different if the game is to be repeated an indefinite number of times. In those circumstances there may be an inducement to cooperation. Robert Axelrod has shown that in a game like the prisoner's dilemma repeated an indefinite number of times, the optimal strategy is *tit-for-tat*—if the other player cooperated on the last play, you cooperate on this play; if she didn't cooperate on the last play, you don't on this play.[14]

---

[13] Can you think of a workable way in which the suspects might have agreed never to confess before they perpetrated the crime? Put in the language of game theory, can a participant in a game like the prisoner's dilemma make a *credible commitment* not to confess if she and her partner are caught?

[14] See ROBERT AXELROD, THE EVOLUTION OF COOPERATION (1984).

These considerations of a fixed versus an indefinite number of plays of a game may seem removed from the concerns of the law, but they really are not. Consider, for example, the relations between a creditor and a debtor. When the debtor's affairs are going well, the credit relations between the creditor and the debtor may be analogized to a game played an indefinite number of times. But if the debtor is likely to become insolvent soon, the relations between debtor and creditor become much more like a game to be played a fixed (and, perhaps, few) number of times. As a result, trust and cooperation between the parties may break down, with the debtor trying to hide his assets and the creditor trying to grab them for resale to recoup his losses.

We shall see that these concepts from game theory will play an important role in our understanding of legal rules and institutions.

VIII. The Theory of Asset Pricing

The area of microeconomic theory that deals with capital and labor markets is beyond the scope of the material in this book. There is, however, one tool from this area that we shall use: the theory of asset pricing.

Assets are resources that generate a stream of income. For instance, an apartment building can generate a stream of rental payments; a patent can generate a stream of royalty payments; an annuity can generate a fixed amount of income per year. There is a technique for converting these various streams of future income (or future expenses or, still more generally, net receipts) into a lump sum today. The general question that is being asked is, "How much would you be prepared to pay today for an asset that generated a given future flow of net receipts in the future?"

We can answer that question by computing what is called the *present discounted value* of the future flow of net receipts. Suppose that ownership of a particular asset will generate F_1 in net receipts at the end of the first year; F_2 in net receipts at the end of the second year; F_3 in net receipts at the end of the third year; and F_n at the end of the nth year. The present discounted value of that asset, supposing that the prevailing rate of interest is r, is equal to:

$$PDV = \frac{F_1}{(1 + r)} + \frac{F_2}{(1 + r)^2} + \frac{F_3}{(1 + r)^3} + \cdots + \frac{F_n}{(1 + r)^n}.$$

This result has many applications to law. For instance, suppose that a court is seeking to compensate someone whose property was destroyed. One method of valuing the loss is to compute the present discounted value of the future flow of net receipts to which the owner was entitled.

IX. General Equilibrium and Welfare Economics

The microeconomic theory we have been reviewing to this point has focused on the fundamental concepts of maximization, equilibrium, and efficiency in describing the decisions of consumers and firms. The part of microeconomic theory called *welfare economics* explores how the decisions of many individuals and firms interact to affect the well-being of individuals as a group. Welfare economics is much more philosophical than other topics in microeconomic theory. Here the great policy issues are raised. For example, is there an inherent conflict between efficiency and fairness? To what extent can unregulated markets maximize individual well-being? When and how should the government intervene in the marketplace? Can economics identify a just distribution of goods and services? In this brief introduction, we can only hint at how microeconomic theory approaches these questions. Nonetheless, this material is fundamental to the economic analysis of legal rules.

A. General Equilibrium and Efficiency Theorems

One of the great accomplishments of modern microeconomics is the specification of the conditions under which the independent decisions of utility-maximizing consumers and profit-maximizing firms will lead to the inevitable, spontaneous establishment of equilibrium in all markets simultaneously. Such a condition is known as *general equilibrium*. General equilibrium will be achieved only when competitive forces have led to the equality of marginal benefit and marginal cost in the market for every single commodity and service. As you can well imagine, this condition is unlikely to be realized in the real world. However, there are two practical reasons for knowing what conditions must hold for general equilibrium to obtain. First, while *all* real-world markets may not obey those conditions, many of them will. Second, the specification of the conditions that lead to general equilibrium provides a benchmark for evaluating various markets and making recommendations for public policy.

Modern microeconomics has demonstrated that general equilibrium has characteristics that economists describe as socially optimal—that is, the general equilibrium is both productively and allocatively efficient.

B. Market Failure

General equilibrium is, in welfare terms, such a desirable outcome that it would be helpful to know the conditions under which it will hold. Stripped of detail, the essential condition is that all markets are perfectly competitive. We can characterize the things that can go wrong to prevent this essential condition from being attained in a market. In this section we shall describe the four sources of *market failure,* as it is called, and describe the public policies that can, in theory, correct those failures.

1. Monopoly and Market Power The first source of market failure is monopoly in its various forms: monopoly in the output market, collusion among otherwise competitive firms or suppliers of inputs, and monopsony (only one buyer) in the input market. If the industry were competitive, marginal benefit and marginal cost would be equal. But as illustrated in Figure 2.11, the monopolist's profit-maximizing output and price combination occurs at a point where the price exceeds the marginal cost of production. The price is too high, and the quantity supplied is too low from the viewpoint of efficiency.

The public policies for correcting the shortcomings of monopoly are to replace monopoly with competition where possible, or to regulate the price charged by the monopolist. The first policy is the rationale for the antitrust laws. But sometimes it is not possible or even desirable to replace a monopoly. Natural monopolies, such as public utilities, are an example; those monopolies are allowed to continue in existence, but government regulates their prices.

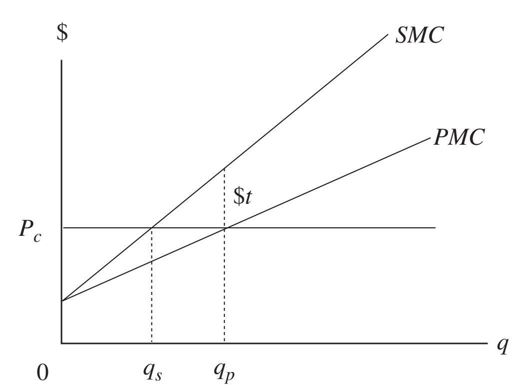
2. Externalities The second source of market failure is the presence of what economists call *externalities*. Exchange inside a market is voluntary and mutually beneficial. Typically, the parties to the exchange capture all the benefits and bear all the costs, thus having the best information about the desirability of the exchange. But sometimes the benefits of an exchange may spill over onto other parties than those explicitly engaged in the exchange. Moreover, the costs of the exchange may also spill over onto other parties. The first instance is an example of an *external benefit*; the second, an *external cost*. An example of an external benefit is the pollination that a beekeeper provides to his neighbor who runs an apple orchard. An example of an external cost is air or water pollution.

Let's explore the idea of an external cost (frequently called simply an *externality*) to see how it can lead to market failure and what public policies can correct this failing. Suppose that a factory located upstream from a populous city dumps toxic materials into the river as a by-product of its production process. This action by the factory imposes an unbargained-for cost on the townspeople downstream: They must incur some additional costs to clean up the water or to bring in safe water from elsewhere. In what way has the market failed in this example? The reason the market fails in the presence of external costs is that the generator of the externality does not have to pay for harming others, and so exercises too little self-restraint. He or she acts as if the cost of disposing of waste is zero, when, in fact, there are real costs involved, as the people downstream can testify. In a technical sense, the externality generator produces too much output and too much harm because there is a difference between *private* marginal cost and *social* marginal cost.

Private marginal cost, in our example, is the marginal cost of production for the factory. Social marginal cost is the sum of private marginal cost and the additional marginal costs involuntarily imposed on third parties by each unit of production. The difference is shown in Figure 2.15. Social marginal cost is greater than private marginal cost at every level of output. The vertical difference between the two curves equals the

FIGURE 2.15

The difference between private and social marginal cost.



amount of the external marginal cost at any level of output. Note that if production is zero, there is no externality, but that as production increases, the amount of external cost per unit of output increases.

The profit-maximizing firm operates along its private marginal cost curve and maximizes profits by choosing that output level for which $P_c = PMC$—namely, q_p. But from society's point of view, this output is too large. Society's resources will be most efficiently used if the firm chooses its output level by equating P_c and SMC at q_s. At that level the firm has taken into account not only its own costs of production but also any costs it imposes on others involuntarily.

What public policies will induce the externality generator to take external costs into account? That is one of the central questions that this book will seek to answer. The key to achieving the social optimum where there are externalities is to induce private profit-maximizers to restrict their output to the socially optimal, not privately optimal, point. This is done by policies that cause the firm to operate along the social marginal cost curve rather than along the private marginal cost curve. When this is accomplished, the externality is said to have been *internalized* in the sense that the private firm now takes it into consideration.

> **QUESTION 2.8:** In Figure 2.15, if the firm is producing output, is there any external cost being generated? If so, why is this output level called a social optimum? Would it not be optimal to have *no* external cost? At what level of output would that occur? Does our earlier discussion that characterized any social optimum as the point at which (social) marginal cost equals (social) marginal benefit provide any guidance? Is the point at which social marginal cost and social marginal benefit are equal consistent with the existence of *some* external cost? Why or why not?

3. Public Goods The third source of market failure is the presence of a commodity called a *public good*. A public good is a commodity with two very closely related characteristics:

1. *Nonrivalrous consumption:* consumption of a public good by one person does not leave less for any other consumer.
2. *Nonexcludability:* the costs of excluding nonpaying beneficiaries who consume the good are so high that no private profit-maximizing firm is willing to supply the good.

Consider national defense. Suppose, for the purposes of illustration, that national defense were provided by competing private companies. For an annual fee a company would sell protection to its customers against loss from foreign invasion by air, land, or sea. Only those customers who purchase some company's services would be protected against foreign invasion. Perhaps these customers could be identified by special garments, and their property denoted by a large white *X* painted on the roof of their homes.

Who will purchase the services of these private national defense companies? Some will but many will not. Many of the nonpurchasers will reason that if their neighbor will purchase a protection policy from a private national defense company, then they, too, will be protected: It will prove virtually impossible for the private company to protect the property and person of the neighbor without also providing security to the nearby nonpurchaser. Thus, the consumption of national defense is nonrivalrous: Consumption by one person does not leave less for any other consumer. For that reason, there is a strong inducement for consumers of the privately provided public good to try to be *free riders*: They hope to benefit at no cost to themselves from the payment of others.

The related problem for the private supplier of a public good is the difficulty of excluding nonpaying beneficiaries. The attempt to distinguish those who have from those who have not subscribed to the private defense companies is almost certain to fail; for example, the identifying clothes and property markings can easily be counterfeited.

As a result of the presence of free riders and the high cost of distinguishing nonpaying from paying beneficiaries, it is not likely that the private company will be able to induce many people to purchase defense services. If private profit-maximizing firms are the only providers of national defense, too little of that good will be provided.

How can public policy correct the market failure in the provision of public goods? There are two general correctives. First, the government may undertake to *subsidize* the private provision of the public good, either directly or indirectly through the tax system. An example might be research on basic science. Second, the government may undertake to provide the public good itself and to pay the costs of providing the service through the revenues raised by compulsory taxation. This is, in fact, how national defense is supplied.

Web Note 2.1

Another kind of problem that markets have is coordinating people, especially when they act collectively. See our website for a discussion of coordination and collective action applied to legal issues.

4. Severe Informational Asymmetries The fourth source of market failure is an imbalance of information between parties to an exchange, one so severe that exchange is impeded.

To illustrate, it is often the case that sellers know more about the quality of goods than do buyers. For example, a person who offers his car for sale knows far more about its quirks than does a potential buyer. Similarly, when a bank presents a depository agreement for the signature of a person opening a checking account, the bank knows far more than the customer about the legal consequences of the agreement.

When sellers know more about a product than do buyers, or vice versa, information is said to be distributed asymmetrically in the market. Under some circumstances, these asymmetries can be corrected by the mechanism of voluntary exchange, for example, by the seller's willingness to provide a warranty to guarantee the quality of a product. But severe asymmetries can disrupt markets so much that a social optimum cannot be achieved by voluntary exchange. When that happens, government intervention in the market can ideally correct for the informational asymmetries and induce more nearly optimal exchange. For example, the purchasers of a home are often at a disadvantage vis-à-vis the current owners in learning of latent defects, such as the presence of termites or a cracked foundation. As a result, the market for the sale of homes may not function efficiently; purchasers may be paying too much for homes or may inefficiently refrain from purchases because of a fear of latent defects. Many states have responded by requiring sellers to disclose knowledge of any latent defects to prospective purchasers of houses. If the sellers do not make this disclosure, then they may be responsible for correcting those defects.

Web Note 2.2

One of the most important issues in welfare economics has been the derivation of a social welfare function, which aggregates individual preferences into social preferences. The Arrow Impossibility Theorem, one of the most significant intellectual achievements of modern economics, argues that a social welfare function with minimally desirable properties cannot be constructed. We describe the theorem in more detail at our website.

C. Potential Pareto Improvements or Kaldor-Hicks Efficiency

Dissatisfied with the Pareto criterion, economists developed the notion of a *potential Pareto improvement* (sometimes called *Kaldor-Hicks efficiency*). This is an attempt to surmount the restriction of the Pareto criterion that only those changes are recommended in which at least one person is made better off and no one is made worse off. That criterion requires that gainers explicitly compensate losers in any change. If there is not explicit payment, losers can veto any change. That is, every change must be by unanimous consent. This has clear disadvantages as a guide to public policy.

By contrast, a potential Pareto improvement allows changes in which there are both gainers and losers but requires that the gainers gain more than the losers lose. If this condition is satisfied, the gainers can, in principle, compensate the losers and still have a surplus left for themselves. For a potential Pareto improvement, compensation does not actually have to be made, but it must be possible in principle. In essence, this is the technique of cost-benefit analysis. In cost-benefit analysis, a project is undertaken when its benefits exceed its costs, which implies that the gainers could compensate the losers. Cost-benefit analysis tries to take into account both the private and social costs and benefits of the action being contemplated. There are both theoretical and empirical problems with this standard, but it is indispensable to applied welfare economics.

Consider how these two criteria—the Pareto criterion and the Kaldor-Hicks criterion—would help us to analyze the efficiency and distributive justice of a manufacturing plant's decision to relocate. Suppose that the plant announces that it is going to move from town *A* to town *B*. There will be gainers—those in town *B* who will be employed by the new plant, the retail merchants and home builders in *B,* the shareholders of the corporation, and so on. But there will also be losers—those in town *A* who are now unemployed, the retail merchants in *A,* the customers of the plant who are now located further away from the plant, and so on. If we were to apply the Pareto criterion to this decision, the gainers would have to pay the losers whatever it would take for them to be indifferent between the plant's staying in A and moving to B. If we were to apply the potential Pareto criterion to this decision, the gainers would have to gain more than the losers lose but no compensation would actually occur.

Web Note 2.3

See our website for much more on cost-benefit analysis as a guide to public policy, including legal change.

X. Decision Making Under Uncertainty: Risk and Insurance

In nearly all of the economic models we have examined so far, we have implicitly assumed that uncertainty did not cloud the decision. This is clearly a simplifying assumption. It is time to expand our basic economic model by explicitly allowing for the presence of uncertainty.

A. Expected Monetary Value

Suppose that an entrepreneur is considering two possible projects in which to invest. The first, D_1, involves the production of an output whose market is familiar and stable. There is no uncertainty about the outcome of project D_1; the entrepreneur can be confident of earning a profit of \$200 if he takes D_1. The second course of action, D_2, involves a novel product whose reception by the consuming public is uncertain. If consumers like the new product, the entrepreneur can earn profits of \$300. However, if they do not like it, he stands to lose \$30.

How is the entrepreneur supposed to compare these two projects? One possibility is to compare their expected monetary values. An *expected value* is the sum of the probabilities of each possible outcome times the value of each of those outcomes. For example, suppose that there are four possible numerical outcomes, labeled O_1 through O_4, to a decision. Suppose also that there are four separate probability estimates, labeled p_1 through p_4, associated with each of the four outcomes. If these are the only possible outcomes, then these probabilities must sum to 1. The expected value (EV) of this decision is then:

$$EV = p_1O_1 + p_2O_2 + p_3O_3 + p_4O_4.$$

To return to our example, the entrepreneur can get \$200 by choosing D_1. What is the expected monetary value of decision D_2? There are two possible outcomes, and in order to perform the calculation the entrepreneur needs to know the probabilities. Let p denote the probability of the new product's succeeding. Thus, $(1 - p)$ is the probability that it fails. Then, the expected monetary value of D_2 for any probability p is given by the expression:

$$EMV(D_2) = 300p + (-30)(1 - p).$$

Thus, if the probability of success for the new product equals $\frac{1}{3}$, the expected monetary value of the decision to introduce that new product equals \$80.

Where does the decision maker get information about the probabilities of the various outcomes? Perhaps the seasoned entrepreneur has some intuition about p or perhaps marketing surveys have provided a scientific basis for assessing p. Still another possibility might be that he calculates the level of p that will make the expected monetary value of D_2 equal to that of the certain event, D_1. A strong reason for doing that would be that, although he might not know for sure what p is, it would be valuable to know how high p must be in order for it to give the same expected profits as the safe course of action, D_1. For example, even if there was no way to know p for sure, suppose that one could calculate that in order for the uncertain course of action to have a higher expected value than the safe course of action, the probability of success of the new product would have to be 0.95, a near certainty. That would be valuable information.

It is a simple matter to calculate the level of p that equates the expected monetary value of D_1 and D_2. That is the p that solves the following equation:

$$300p - 30(1 - p) = 200,$$

which implies that $p = .7$. The implication, of course, is that if the probability of the new product's success is .7 or greater, then D_2 has a higher expected monetary value than does D_1 and the entrepreneur will choose D_2.

B. Maximization of Expected Utility: Attitudes Toward Risk

Do people deal with uncertainty by maximizing expected monetary values? Suppose that the two decisions of the previous section, D_1 and D_2, have the same expected monetary value. Would you be indifferent between the two courses of action? Probably not. D_1 is a sure thing. D_2 is not. Upon reflection, many would hesitate to take D_2 unless the expected monetary value of D_2 was greater than that of D_1. The reason for this hesitation may lie in the fact that many of us are reluctant to gamble, and D_2 certainly is a gamble. We are generally much more comfortable with a sure thing like D_1. Can we formalize our theory of decision making under uncertainty to take account of this attitude?

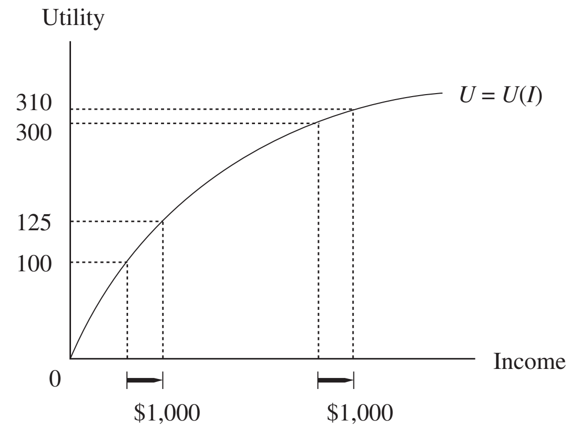
The formal explanation for this phenomenon of avoiding gambles was first offered in the eighteenth century by the Swiss mathematician and cleric Daniel Bernoulli. Bernoulli often noticed that people who make decisions under uncertainty do not attempt to maximize expected monetary values. Rather, they maximize expected utility. The introduction of utility allows us to introduce the notion of decision makers' attitudes toward risk.

1. Risk Aversion

Assume that utility is a function of, among other things, money income:

$$U = U(I).$$

FIGURE 2.16
Risk aversion as diminishing marginal utility of income.



Bernoulli suggested that a common relationship between money income and utility was that as income increased, utility also increased, but at a decreasing rate. Such a utility function exhibits diminishing marginal utility of income. For example, if one's income level is $10,000, an additional $100 in income will add more to one's total utility than will $100 added to that same person's income of $40,000. A utility function like that shown in Figure 2.16 has this property. When this person's income is increased by $1000 at a low level of income, her utility increases from 100 to 125 units, an increase of 25 units. But when her income is increased by $1000 at a higher level of income, her utility increases from 300 to 310 units, an increase of only 10 units.

A person who has diminishing marginal utility from money income is said to be *risk-averse.* Here is a more formal definition of risk aversion:

> A person is said to be risk-averse if she considers the utility of a certain prospect of money income to be higher than the expected utility of an uncertain prospect of equal expected monetary value.

For example, in the preceding entrepreneur's project, a risk-averse decision maker might prefer to have $80 for certain rather than undertake a project whose *EMV* equals $80.

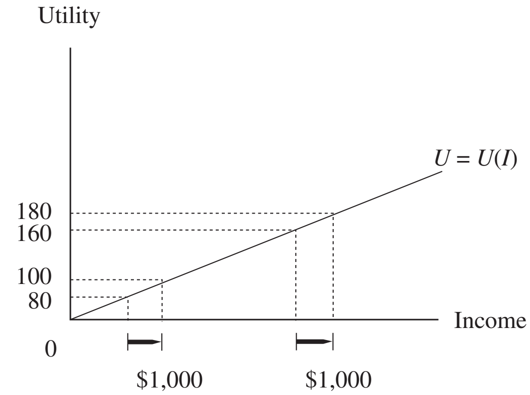
2. Risk Neutrality Economists presume that most people are averse toward risk, but some people are either neutral toward risk or, like gamblers, rock climbers, and race car drivers, prefer risk. Like aversion, these attitudes toward risk may also be defined in terms of the individual's utility function in money income and the marginal utility of income.

Someone who is *risk neutral* has a constant marginal utility of income and is, therefore, indifferent between a certain prospect of income and an uncertain prospect of equal expected monetary value. Figure 2.17 gives the utility function for a risk-neutral person. It is a straight line because the marginal utility of income to a risk-neutral person is constant.

The figure compares the change in utility when the risk-neutral person's income is increased by $1000 at two different levels of income. When this person's income is

FIGURE 2.17

Risk neutrality as constant marginal utility of income.



increased by $1000 at a low level of income, his utility increases from 80 to 100 units, an increase of 20 units. And when his income is increased by $1000 at a high level of income, his utility increases by exactly the same amount, 20 units, from 160 to 180 units. Thus, for the risk-neutral person the marginal utility of income is constant.

Economists and finance specialists very rarely attribute an attitude of risk-neutrality to individuals. However, they quite commonly assume that business organizations are risk-neutral.

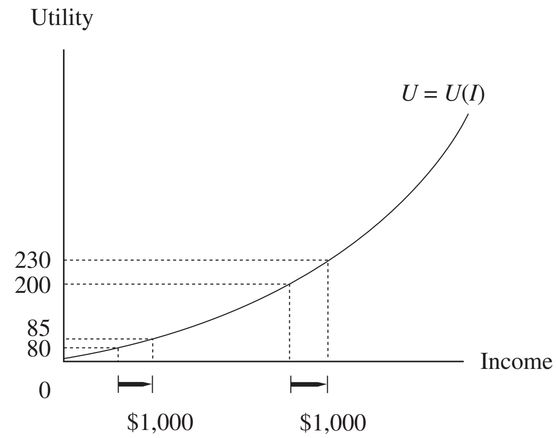
3. Risk-Seeking or Risk-Preferring Someone who is *risk-seeking* or *risk-preferring* has an increasing marginal utility of income and, therefore, prefers an uncertain prospect of income to a certain prospect of equal expected monetary value. Figure 2.18 gives the utility function of a risk-preferring individual. The figure allows us to compare the change in utility when the risk-preferring person's income is increased by $1000 at two different levels of income. When this person's income is increased by $1000 at a low level of income, her utility increases from 80 to 85 units, an increase of 5 units. However, when her income is increased by $1000 at a high level of income, her utility increases from 200 to 230 units, an increase of 30 units. Thus, for the risk-preferring person the marginal utility of income increases.

Web Note 2.4

One of the winners of the Nobel Prize in Economics in 2002 was Daniel Kahneman, Professor Emeritus of Psychology at Princeton University. Kahneman and his coauthor, the late Amos Tversky, did experiments to see the extent to which people's attitudes toward risk fit those we have just studied. The experiments suggested that most people have complex feelings about losses and gains that Kahneman and Tversky characterized as "loss aversion." See section XII, on page 50 and our website for more on the experiments and their implications.

FIGURE 2.18

Risk preferring as increasing marginal utility of income.



C. The Demand for Insurance

One of the most important behavioral implications of risk aversion is that people will pay money to avoid having to face uncertain outcomes. That is, a risk-averse person might prefer a lower certain income to a higher uncertain income.

There are three ways in which a risk-averse person may convert an uncertain into a certain outcome. First, he may purchase insurance from someone else. In exchange for giving up a certain amount of income (the insurance premium), the insurance company will bear the risk of the uncertain event. The risk-averse person considers himself better off with the lower certain income than facing the uncertain higher income. Second, he may self-insure. This may involve incurring expenses to minimize the probability of an uncertain event's occurring or to minimize the monetary loss in the event of a particular contingency. An example is the installation of smoke detectors in a home. Another form of self-insurance is the setting aside of a sum of money to cover possible losses. Third, a risk-averse person who is considering the purchase of some risky asset may reduce the price he is willing to pay for that asset.

D. The Supply of Insurance

The material of the previous section concerns the *demand* for insurance by risk-averse individuals. Let us now turn to a brief consideration of the supply of insurance by profit-maximizing insurance companies.

Insurance companies are presumed to be profit-maximizing firms. They offer insurance contracts not because they prefer gambles to certainties but because of a mathematical theorem known as the *law of large numbers*. This law holds that unpredictable events for individuals become predictable among large groups of individuals. For example, none of us knows whether our house will burn down next year. But the occurrence of fire in a city, state, or nation is regular enough so that an insurance company can easily determine the objective probabilities. By insuring a large number of people, an insurance company can predict the total amount of claims.

1. Moral Hazard *Moral hazard* arises when the behavior of the insured person or entity changes after the purchase of insurance so that the probability of loss or the size of the loss increases. An extreme example is the insured's incentive to burn his home when he has been allowed to insure it for more than its market value. A more realistic example comes from loss due to theft. Suppose that you have just purchased a new sound system for your car but that you do not have insurance to cover your loss from theft. Under these circumstances you are likely to lock your car whenever you leave it, to park it in well-lighted places at night, to patronize only well-patrolled parking garages, and so on.

Now suppose that you purchase an insurance policy that will, unrealistically, compensate you for the full cost of any insured loss that you suffer. With the policy in force you now may be less assiduous about locking your car or parking in well-lighted places. In short, the very fact that your loss is insured may cause you to act so as to increase the probability of a loss.

Insurance companies attempt to set their premiums so that, roughly, the premium modestly exceeds the expected monetary value of the loss. Therefore, a premium that has been set without regard for the increased probability of loss due to moral hazard will be too low and thus threaten the continued profitability of the firm. Every insurer is aware of this problem and has developed methods to minimize it. Among the most common are *coinsurance* and *deductibles*. Under coinsurance the insuree shoulders a fixed percentage of his loss, with the insurer picking up the remaining portion of the loss; under a deductible plan, the insuree shoulders a fixed dollar amount of the loss, with the insurance company paying for all losses above that amount. In addition, some insurance companies offer reductions in premiums for certain easily established acts that reduce claims. For example, life and health insurance premiums are less for nonsmokers; auto insurance premiums are less for nondrinkers; and fire insurance rates are lower for those who install smoke detectors.

2. Adverse Selection The other major problem faced by insurance companies is called *adverse selection*. This arises because of the high cost to insurers of accurately distinguishing between high- and low-risk insurees. Although the law of large numbers helps the company in assessing probabilities, what it calculates from the large sample are average probabilities. The insurance premium must be set using this average probability of a particular loss. For example, insurance companies have determined that unmarried males between the ages of 16 and, say, 25, have a much higher likelihood of being in an automobile accident than do other identifiable groups of drivers. As a result, the insurance premium charged to members of this group is higher than that charged to other groups whose likelihood of accident is much lower.

But even though unmarried males between the ages of 16 and 25 are, on average, much more likely to be involved in an accident, there are some young men within that group who are even more reckless than average and some who are much less reckless than the group's average. If it is difficult for the insurer to distinguish these groups from the larger group of unmarried males aged 16 to 25, then the premium that is set equal to the average likelihood of harm within the group will seem like a bargain to those who know they are reckless and too high to those who know that they are safer than their peers.

Let us assume, as seems reasonable, that in many cases the individuals know better than the insurance company what their true risks are. For example, the insured alone may know that he drinks heavily and smokes in bed or that he is intending to murder his spouse, in whose insurance policy he has just been named principal beneficiary. If so, then this asymmetrical information may induce only high-risk people to purchase insurance and low-risk people to purchase none. As a result, the insurance company will find itself paying out more in claims than it had anticipated. It may, therefore, raise the premiums higher. This will drive out some more relatively safer customers, leaving an even-riskier clientele behind. The incidence and volume of claims may go up yet again, setting off a new round of premium increases, defections of less-risky customers, and so on. This process is referred to as an insurance *death spiral*.

The same devices that insurance companies employ to minimize risks of moral hazards also may serve to minimize the adverse selection problem. Coinsurance and deductible provisions are much less attractive to high-risk than to low-risk insurees so that an insuree's willingness to accept those provisions may indicate to the insurance company to which risk class the applicant belongs. Exclusion of benefits for loss arising from preexisting conditions is another method of trying to distinguish high- and low-risk people. The insurer can also attempt, over a longer time horizon, to reduce the adverse selection bias by developing better methods of discriminating among the insured, such as medical and psychological testing, so as to place insurees in more accurate risk classes. Finally, insurers frequently practice *experience rating*—the practice of adjusting the insuree's premium up or down according to his experience of insurable losses. If an insuree appears to be accident prone, then the insurer may raise his premium to reflect the greater probability or size of loss. In the limit, the insurer may refuse to cover the insuree.

XI. Profits and Growth[15]

Imagine a banker who asks to be paid by placing one penny on the first square of a chess board, two pennies on the second square, four on the third, and so on. Using only the white squares, the initial penny would double in value thirty-one times, leaving $21.5 million on the last white square. Growth compounds faster than the mind can grasp. In 1900 Argentina's income per person resembled Canada's, and today Canada's is more than three times higher. After World War II, Korea and Nigeria had similar national income per person, and today Korea's is nineteen times higher. Most people cannot imagine China with more economic influence in the world than the United States, but, if current trends continue, China will surpass the United States in national income in 2014.[16] Lifting so many people out of poverty in East Asia in the late twentieth century is one of history's remarkable accomplishments. In contrast, one of history's depressing economic failures in the late twentieth century is sub-Saharan Africa, where GDP per person declined since 1975 roughly by 25 percent.

Why do some countries grow faster than others? Sustained growth requires innovation. An innovation occurs when someone discovers a better way to make something or something better to make. Entrepreneurs make things in better ways by improving organizations

[15] This section draws on Chapter 1 of ROBERT COOTER & HANS-BERND SCHAEFER, LAW AND THE POVERTY OF NATIONS (2011).

[16] Because China's population is 4 to 5 times greater than the United States' population, China's income per capita in 2014 will still be one-fourth to one-fifth that of the United States. This prediction was made by Carl J. Dahlman, Luce Professor of International Affairs and Information, Georgetown University.

and markets, and scientists invent better things to make. Growth will remain mysterious until economics has an adequate theory of innovation. The only contribution to growth theory so far that merited a Nobel Prize in Economics shows the consequences of innovation for capital and labor but does not attempt to explain innovation.[17]

Law, we believe, is part of the mystery's solution. When an innovator has a new idea, it must be developed in order for the economy to grow. Combining new ideas and capital runs into a fundamental obstacle illustrated by this example: An economist who worked at a Boston investment bank received a letter that read, "I know how your bank can make $10 million. If you give me $1 million, I will tell you." The letter captures concisely the problem of financing innovation: The bank does not want to pay for information without first determining its worth, and the innovator fears disclosing information to the bank without first getting paid. Law is central to solving this problem. Later chapters in this book mention "transactional lawyers," who use law to overcome the mistrust that prevents people from cooperating in business. The most fundamental bodies of transaction law are property and contracts, which we cover in Chapters 4, 5, 8, and 9. Making these bodies of law efficient promotes economic growth by uniting innovative ideas and capital. Countries with efficient property and contract have established the legal foundation for innovation and growth.

XII. Behavioral Economics

In our review of microeconomic theory we have followed modern microeconomists in assuming that decision makers are rationally self-interested. This theory of decision making is called *rational choice theory* and has served the economics profession well over the past 60 or more years in theorizing about how people make explicitly economic decisions. But rational choice theory has been under attack over the past 30 years or so. This attack has been principally empirical. That is, it has been premised on experimental findings that people do not behave in the ways predicted by rational choice theory. (We will give examples shortly.)

Two principal names in the experimental literature that have been critical of rational choice theory are Daniel Kahneman and Amos Tversky. As we noted above, Kahneman won the Nobel Prize in Economics in 2002. The body of literature inspired by Kahneman and Tversky has as acquired the name *behavioral economics* and the legal analysis that takes account of these findings is called *behavioral law and economics*.[18]

What are some examples of behavioral economics? Let us consider two (although there are many others) that have particular relevance to the law. The first is the result of experiments involving the *ultimatum bargaining game*. In that game there are two players, neither of whom knows the identity of the other. They interact anonymously. Their task in the game is to divide a small sum of money—say, $20. One player is designated

---

[17] In 1987 Robert Solow received the Nobel Prize in Economics for his contributions to economic growth theory.

[18] For a summary of the field, see Russell B. Korobkin & Thomas S. Ulen, *Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics*, 88 CAL. L. REV. 1051 (2000).

Player 1; the other, Player 2. Player 1 makes some proposal to Player 2 about how the $20 should be divided; Player 2 can then either accept that proposal, in which case the experimenter actually gives the two players the amounts proposed by Player 1, or reject the proposal, in which case neither player gets any money.

Rational choice theory predicts that Player 1 will take advantage of her position as the proposer to give herself a disproportionate share of the $20—say, $15, leaving $5 for Player 2. Player 1 might also reason that Player 2 will take $5 rather than nothing. And, in fact, Player 2 might think that Player 1 is selfish, but that, after all, $5 is better than nothing.

This game has been played in experiments in over 140 countries and among groups with vastly different incomes, ages, education levels, religions, and the like, and in countries that are wealthy and countries that are poor. The modal (most common) outcome is a 50-50 split of the stakes—each player gets $10. Equally interesting is the fact that in many countries if Player 1 tries to take more than 70 percent of the stakes (more than $14 in our $20 example), Player 2 rejects the offer and they each get nothing.[19]

We take some heart from the fact that strangers seldom take advantage of one another in the ultimatum bargaining game. Rather, the norm seems to be to treat the other party fairly; in fact, to treat him or her exactly as one treats oneself. Are there legal implications of this insight? We return to this point in Chapters 8 and 9 when we consider the concern of contract law with some forms of advantage-taking in bargaining.

A second example of a behavioral economics finding is the *hindsight bias*. This refers to the fact that things that actually happen seem, in hindsight (*ex post*), to have been far more likely than they were in foresight (*ex ante*). So, if you asked people in spring 2010 the probability that Spain would win the 2010 World Cup in South Africa, they might have said, "10 percent." But if you ask them after Spain actually won the cup, they will say it was much more likely—say, 40 percent. Is there a legal implication? Consider something that we will discuss in Chapter 6: How do we induce someone to take the right amount of precaution against harming another person? As we will see, one way to do that is to expose the possible injurer to liability for the injuries that a victim suffers if the injurer did not take a prudent amount of precaution. Here's the problem—what might seem prudent precaution *before* an accident occurs might appear, in hindsight, to have been imprudent. That is, if an accident has occurred, the hindsight bias may tell us that the accident was more inevitable than we would have thought before.

Rational choice theory cannot explain the observed behavior in the ultimatum bargaining game or hindsight bias. The central insight of behavioral economics is that human beings make predictable errors in judgment, cognition, and decision making. They are, to quote the title of a book on this topic by Dan Ariely, "predictably irrational." Economic analysis should use rational choice theory or behavioral theory, depending on which one predicts the law's effects on the behavior more accurately.

[19] The group that typically plays this game in line with the predictions of rational choice theory (in which Player 1 makes a proposal to take much more than half of the stakes and Player 2 accepts that) are graduate students in economics. Did they select a graduate study in economics because they already find rational choice theory attractive? Or did their graduate studies in economics convince them that rational choice was the appropriate way to behave?