profits in
.bor costs.

CHAPTER 25

MONOPOLY

In the preceding chapters we have analyzed the behavior of a competitive industry, a market structure that is most likely when there are a large number of small firms. In this chapter we turn to the opposite extreme and consider an industry structure when there is only *one* firm in the industry—a **monopoly**.

When there is only one firm in a market, that firm is very unlikely to take the market price as given. Instead, a monopoly would recognize its influence over the market price and choose that level of price and output that maximized its overall profits.

Of course, it can't choose price and output independently; for any given price, the monopoly will be able to sell only what the market will bear. If it chooses a high price, it will be able to sell only a small quantity. The demand behavior of the consumers will constrain the monopolist's choice of price and quantity.

We can view the monopolist as choosing the price and letting the consumers choose how much they wish to buy at that price, or we can think of the monopolist as choosing the quantity, and letting the consumers decide what price they will pay for that quantity. The first approach is probably more natural, but the second turns out to be analytically more convenient. Of course, both approaches are equivalent when done correctly.

25.1 Maximizing Profits

We begin by studying the monopolist's profit-maximization problem. Let us use $p(y)$ to denote the market inverse demand curve and $c(y)$ to denote the cost function. Let $r(y) = p(y)y$ denote the revenue function of the monopolist. The monopolist's profit-maximization problem then takes the form

$$\max_{y} \; r(y) - c(y).$$

The optimality condition for this problem is straightforward: at the optimal choice of output we must have marginal revenue equal to marginal cost. If marginal revenue were less than marginal cost it would pay the firm to decrease output, since the savings in cost would more than make up for the loss in revenue. If the marginal revenue were greater than the marginal cost, it would pay the firm to increase output. The only point where the firm has no incentive to change output is where marginal revenue equals marginal cost.

In terms of algebra, we can write the optimization condition as

$$MR = MC$$

or

$$\frac{\Delta r}{\Delta y} = \frac{\Delta c}{\Delta y}.$$

The same $MR = MC$ condition has to hold in the case of a competitive firm; in that case, marginal revenue is equal to the price and the condition reduces to price equals marginal cost.

In the case of a monopolist, the marginal revenue term is slightly more complicated. If the monopolist decides to increase its output by Δy, there are two effects on revenues. First it sells more output and receives a revenue of $p\Delta y$ from that. But second, the monopolist pushes the price down by Δp and it gets this lower price on *all* the output it has been selling.

Thus the total effect on revenues of changing output by Δy will be

$$\Delta r = p\Delta y + y\Delta p,$$

so that the change in revenue divided by the change in output—the marginal revenue—is

$$\frac{\Delta r}{\Delta y} = p + \frac{\Delta p}{\Delta y}y.$$

(This is exactly the same derivation we went through in our discussion of marginal revenue in Chapter 15. You might want to review that material before proceeding.)

Another way to think about this is to think of the monopolist as choosing its output and price simultaneously—recognizing, of course, the constraint imposed by the demand curve. If the monopolist wants to sell more output it has to lower its price. But this lower price will mean a lower price for all of the units it is selling, not just the new units. Hence the term $y\Delta p$.

In the competitive case, a firm that could lower its price below the price charged by other firms would immediately capture the entire market from its competitors. But in the monopolistic case, the monopoly already has the entire market; when it lowers its price, it has to take into account the effect of the price reduction on all the units it sells.

Following the discussion in Chapter 15, we can also express marginal revenue in terms of elasticity via the formula

$$MR(y) = p(y)\left[1 + \frac{1}{\epsilon(y)}\right]$$

and write the "marginal revenue equals marginal costs" optimality condition as

$$p(y)\left[1 + \frac{1}{\epsilon(y)}\right] = MC(y). \qquad (25.1)$$

Since elasticity is naturally negative, we could also write this expression as

$$p(y)\left[1 - \frac{1}{|\epsilon(y)|}\right] = MC(y).$$

From these equations it is easy to see the connection with the competitive case: in the competitive case, the firm faces a flat demand curve—an infinitely elastic demand curve. This means that $1/|\epsilon| = 1/\infty = 0$, so the appropriate version of this equation for a competitive firm is simply price equals marginal cost.

Note that a monopolist will never choose to operate where the demand curve is *inelastic*. For if $|\epsilon| < 1$, then $1/|\epsilon| > 1$, and the marginal revenue is negative, so it can't possibly equal marginal cost. The meaning of this becomes clear when we think of what is implied by an inelastic demand curve: if $|\epsilon| < 1$, then reducing output will increase revenues, and reducing output must reduce total cost, so profits will necessarily increase. Thus any point where $|\epsilon| < 1$ cannot be a profit maximum for a monopolist, since it could increase its profits by producing less output. It follows that a point that yields maximum profits can only occur where $|\epsilon| \geq 1$.

25.2 Linear Demand Curve and Monopoly

Suppose that the monopolist faces a linear demand curve

$$p(y) = a - by.$$

Then the revenue function is

$$r(y) = p(y)y = ay - by^2,$$

and the marginal revenue function is

$$MR(y) = a - 2by.$$

(This follows from the formula given at the end of Chapter 15. It is easy to derive using simple calculus. If you don't know calculus, just memorize the formula, since we will use it quite a bit.)

Note that the marginal revenue function has the same vertical intercept, a, as the demand curve, but it is twice as steep. This gives us an easy way to draw the marginal revenue curve. We know that the vertical intercept is a. To get the horizontal intercept, just take half of the horizontal intercept of the demand curve. Then connect the two intercepts with a straight line. We have illustrated the demand curve and the marginal revenue curve in Figure 25.1.

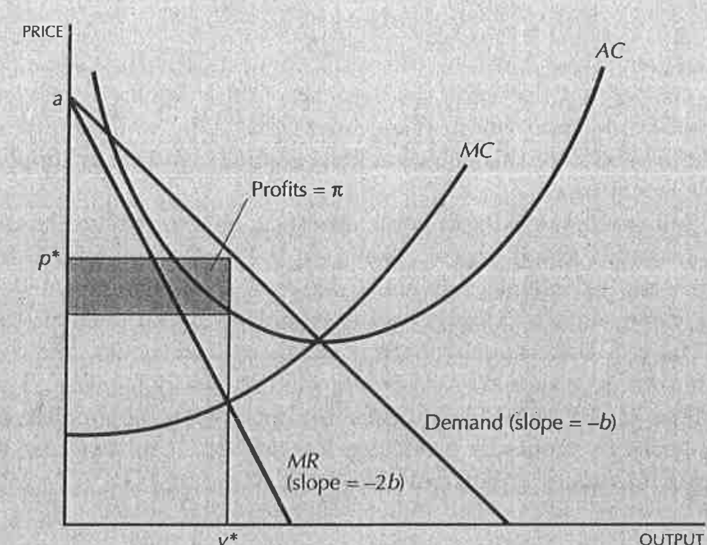

Figure 25.1

Monopoly with a linear demand curve. The monopolist's profit-maximizing output occurs where marginal revenue equals marginal cost.

The optimal output, y^*, is where the marginal revenue curve intersects the marginal cost curve. The monopolist will then charge the maximum price it can get at this output, $p(y^*)$. This gives the monopolist a revenue of $p(y^*)y^*$ from which we subtract the total cost $c(y^*) = AC(y^*)y^*$, leaving a profit area as illustrated.

25.3 Markup Pricing

We can use the elasticity formula for the monopolist to express its optimal pricing policy in another way. Rearranging equation (25.1) we have

$$p(y) = \frac{MC(y^*)}{1 - 1/|\epsilon(y)|}. \tag{25.2}$$

This formulation indicates that the market price is a markup over marginal cost, where the amount of the markup depends on the elasticity of demand. The markup is given by

$$\frac{1}{1 - 1/|\epsilon(y)|}.$$

Since the monopolist always operates where the demand curve is elastic, we are assured that $|\epsilon| > 1$, and thus the markup is greater than 1.

In the case of a constant-elasticity demand curve, this formula is especially simple since $\epsilon(y)$ is a constant. A monopolist who faces a constant-elasticity demand curve will charge a price that is a *constant* markup on marginal cost. This is illustrated in Figure 25.2. The curve labeled $MC/(1 - 1/|\epsilon|)$ is a constant fraction higher than the marginal cost curve; the optimal level of output occurs where $p = MC/(1 - 1/|\epsilon|)$.

EXAMPLE: The Impact of Taxes on a Monopolist

Let us consider a firm with constant marginal costs and ask what happens to the price charged when a quantity tax is imposed. Clearly the marginal costs go up by the amount of the tax, but what happens to the market price?

Let's first consider the case of a linear demand curve, as depicted in Figure 25.3. When the marginal cost curve, MC, shifts up by the amount of the tax to $MC+t$, the intersection of marginal revenue and marginal cost moves to the left. Since the demand curve is half as steep as the marginal revenue curve, the price goes up by half the amount of the tax.

This is easy to see algebraically. The marginal revenue equals marginal cost plus the tax condition is

$$a - 2by = c + t.$$

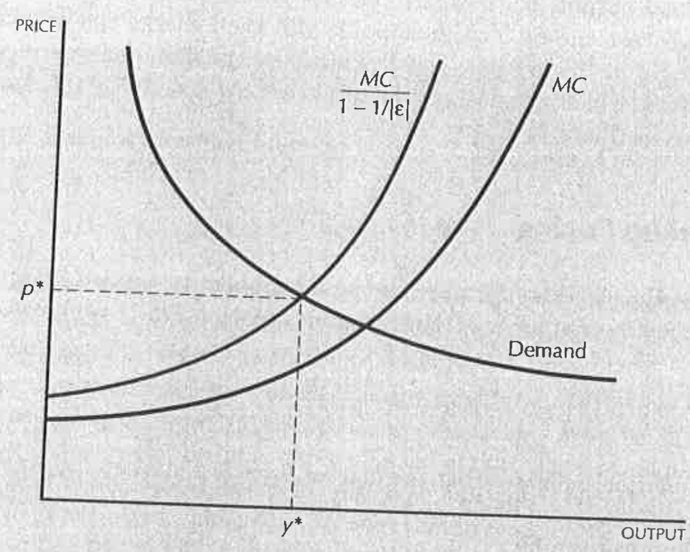

Figure 25.2

Monopoly with constant elasticity demand. To locate the profit-maximizing output level we find the output level where the curve $MC/(1 - 1/|\epsilon|)$ crosses the demand curve.

Solving for y yields

$$y = \frac{a - c - t}{2b}.$$

Thus the change in output is given by

$$\frac{\Delta y}{\Delta t} = -\frac{1}{2b}.$$

The demand curve is

$$p(y) = a - by,$$

so price will change by $-b$ times the change in output:

$$\frac{\Delta p}{\Delta t} = -b \times -\frac{1}{2b} = \frac{1}{2}.$$

In this calculation the factor 1/2 occurs because of the assumptions of the linear demand curve and constant marginal costs. Together these assumptions imply that the price rises by less than the tax increase. Is this likely to be true in general?

The answer is no—in general a tax may increase the price by more or less than the amount of the tax. For an easy example, consider the case of a monopolist facing a constant-elasticity demand curve. Then we have

$$p = \frac{c + t}{1 - 1/|\epsilon|},$$

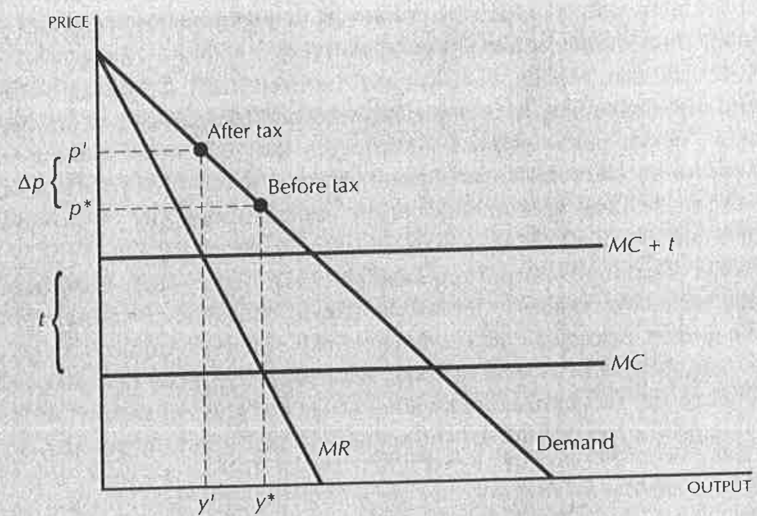

Figure 25.3

Linear demand and taxation. Imposition of a tax on a monopolist facing a linear demand. Note that the price will rise by half the amount of the tax.

so that

$$\frac{\Delta p}{\Delta t} = \frac{1}{1 - 1/|\epsilon|},$$

which is certainly bigger than 1. In this case, the monopolist passes on *more* than the amount of the tax.

Another kind of tax that we might consider is the case of a profits tax. In this case the monopolist is required to pay some fraction τ of its profits to the government. The maximization problem that it faces is then

$$\max_y \; (1 - \tau)[p(y)y - c(y)].$$

But the value of y that maximizes profits will also maximize $(1 - \tau)$ times profits. Thus a pure profits tax will have no effect on a monopolist's choice of output.

25.4 Inefficiency of Monopoly

A competitive industry operates at a point where price equals marginal cost. A monopolized industry operates where price is greater than marginal cost. Thus in general the price will be higher and the output lower if a firm behaves monopolistically rather than competitively. For this reason, consumers will typically be worse off in an industry organized as a monopoly than in one organized competitively.

But, by the same token, the firm will be better off! Counting both the firm and the consumer, it is not clear whether competition or monopoly will be a "better" arrangement. It appears that one must make a value judgment about the relative welfare of consumers and the owners of firms. However, we will see that one can argue against monopoly on grounds of efficiency alone.

Consider a monopoly situation, as depicted in Figure 25.4. Suppose that we could somehow costlessly force this firm to behave as a competitor and take the market price as being set exogenously. Then we would have (p_c, y_c) for the competitive price and output. Alternatively, if the firm recognized its influence on the market price and chose its level of output so as to maximize profits, we would see the monopoly price and output (p_m, y_m).

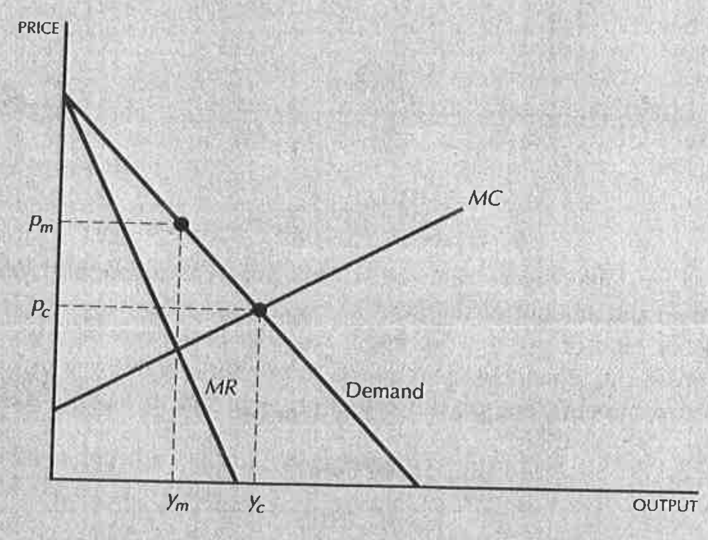

Figure 25.4 **Inefficiency of monopoly.** A monopolist produces less than the competitive amount of output and is therefore Pareto inefficient.

Recall that an economic arrangement is Pareto efficient if there is no way to make anyone better off without making somebody else worse off. Is the monopoly level of output Pareto efficient?

Remember the definition of the inverse demand curve. At each level of output, $p(y)$ measures how much people are willing to pay for an additional unit of the good. Since $p(y)$ is greater than $MC(y)$ for all the output levels between y_m and y_c, there is a whole range of output where people are willing to pay more for a unit of output than it costs to produce it. Clearly there is a potential for Pareto improvement here!

For example, consider the situation at the monopoly level of output y_m. Since $p(y_m) > MC(y_m)$ we know that there is someone who is willing to pay more for an extra unit of output than it costs to produce that extra unit. Suppose that the firm produces this extra output and sells it to this person at any price p where $p(y_m) > p > MC(y_m)$. Then this consumer is made better off because he or she was just willing to pay $p(y_m)$ for that unit of consumption, and it was sold for $p < p(y_m)$. Similarly, it cost the monopolist $MC(y_m)$ to produce that extra unit of output and it sold it for $p > MC(y_m)$. All the other units of output are being sold for the same price as before, so nothing has changed there. But in the sale of the extra unit of output, each side of the market gets some extra surplus—each side of the market is made better off and no one else is made worse off. We have found a Pareto improvement.

It is worthwhile considering the reason for this inefficiency. The efficient level of output is when the willingness to pay for an extra unit of output just equals the cost of producing this extra unit. A competitive firm makes this comparison. But a monopolist also looks at the effect of increasing output on the revenue received from the **inframarginal** units, and these inframarginal units have nothing to do with efficiency. A monopolist would always be ready to sell an additional unit at a lower price than it is currently charging if it did not have to lower the price of all the other inframarginal units that it is currently selling.

25.5 Deadweight Loss of Monopoly

Now that we know that a monopoly is inefficient, we might want to know just how inefficient it is. Is there a way to measure the total loss in efficiency due to a monopoly? We know how to measure the loss to the consumers from having to pay p_m rather than p_c—we just look at the change in consumers' surplus. Similarly, for the firm we know how to measure the gain in profits from charging p_m rather than p_c—we just use the change in producer's surplus.

The most natural way to combine these two numbers is to treat the firm—or, more properly, the owners of the firm—and the consumers of the firm's output symmetrically and add together the profits of the firm and the consumers' surplus. The change in the profits of the firm—the change in producer's surplus—measures how much the owners would be willing to pay to get the higher price under monopoly, and the change in consumers' surplus measures how much the consumers would have to be paid to compensate them for the higher price. Thus the difference between these two numbers should give a sensible measure of the net benefit or cost of the monopoly.

The changes in the producer's and consumers' surplus from a movement from monopolistic to competitive output are illustrated in Figure 25.5. The monopolist's surplus goes down by A due to the lower price on the units he was already selling. It goes up by C due to the profits on the extra units it is now selling.

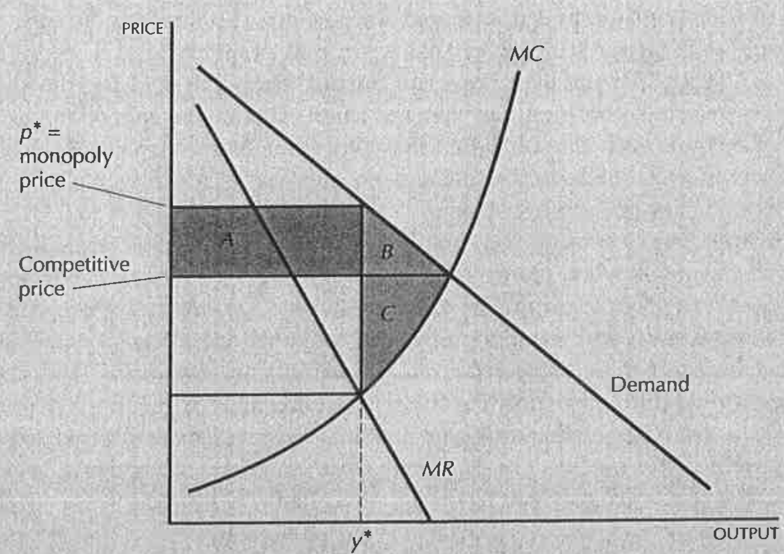

Figure 25.5 **Deadweight loss of monopoly.** The deadweight loss due to the monopoly is given by the area $B + C$.

The consumers' surplus goes up by A, since the consumers are now getting all the units they were buying before at a cheaper price; and it goes up by B, since they get some surplus on the extra units that are being sold. The area A is just a transfer from the monopolist to the consumer; one side of the market is made better off and one side is made worse off, but the total surplus doesn't change. The area $B + C$ represents a true increase in surplus—this area measures the value that the consumers and the producers place on the extra output that has been produced.

The area $B + C$ is known as the **deadweight loss** due to the monopoly. It provides a measure of how much worse off people are paying the monopoly price than paying the competitive price. The deadweight loss due to monopoly, like the deadweight loss due to a tax, measures the value of the lost output by valuing each unit of lost output at the price that people are willing to pay for that unit.

To see that the deadweight loss measures the value of the lost output, think about starting at the monopoly point and providing one additinal unit of output. The value of that marginal unit of output is the market price. The cost of producing the additional unit of output is the marginal cost. Thus the "social value" of producing an extra unit will be simply the price minus the marginal cost. Now consider the value of the next unit of output; again its social value will be the gap between price and marginal cost at that level of output. And so it goes. As we move from the monopoly level of output to the competitive level of output, we "sum up" the distances between the demand curve and the marginal cost curve to generate the value of the lost output due to the monopoly behavior. The total area between the two curves from the monopoly output to the competitive output is the deadweight loss.

EXAMPLE: The Optimal Life of a Patent

A **patent** offers inventors the exclusive right to benefit from their inventions for a limited period of time. Thus a patent offers a kind of limited monopoly. The reason for offering such patent protection is to encourage innovation. In the absence of a patent system, it is likely that individuals and firms would be unwilling to invest much in research and development, since any new discoveries that they would make could be copied by competitors.

In the United States the life of a patent is 17 years. During that period, the holders of the patent have a monopoly on the invention; after the patent expires, anyone is free to utilize the technology described in the patent. The longer the life of a patent, the more gains can be accrued by the inventors, and thus the more incentive they have to invest in research and development. However, the longer the monopoly is allowed to exist, the more deadweight loss will be generated. The benefit from a long patent life is that it encourages innovation; the cost is that it encourages monopoly. The "optimal" patent life is the period that balances these two conflicting effects.

The problem of determining the optimal patent life has been examined by William Nordhaus of Yale University.[1] As Nordhaus indicates, the problem is very complex and there are many unknown relationships involved. Nevertheless, some simple calculations can give some insight as to whether

[1] William Nordhaus, *Invention, Growth, and Welfare* (Cambridge, Mass.: M.I.T. Press, 1969).

the current patent life is wildly out of line with the estimated benefits and costs described above.

Nordhaus found that for "run-of-the-mill" inventions, a patent life of 17 years was roughly 90 percent efficient—meaning that it achieved 90 percent of the maximum possible consumers' surplus. On the basis of these figures, it does not seem like there is a compelling reason to make drastic changes in the patent system.

EXAMPLE: Patent Thickets

The intellectual property protection offered by patents provides incentives to innovate, but this right can be abused. Some observers have argued that the extensions of intellectual property rights to business processes, software, and other domains has resulted in lower patent quality.

One might think of patents as having three dimensions: length, width, and height. The "length" is the time that the patent protection applies. The "width" is how broadly the claims in the patent are interpreted. The "height" is the standard of novelty applied in determining whether the patent really represents a new idea. Unfortunately, only the length is easily quantified. The other aspects of patent quality, breadth, and novelty, can be quite subjective.

Since it has become so easy to acquire patents in recent years, many firms have invested in acquiring patent portfolios on nearly every aspect of their business. Any company that wants to enter a business and compete with an incumbent who owns a broad range of patents may find itself encumbered in a **patent thicket**.

Even firms that are already well established find it important to invest in acquiring a patent portfolio. In 2004, Microsoft paid $440 million to InterTrust Technology to license a portfolio of patents related to computer security, and signed a 10-year pact with Sun Microsystems in which it paid $900 million to resolve patent issues. During 2003–04, Microsoft was granted over 1,000 patents.

Why the emphasis on patent portfolios? For large companies like Microsoft, their primary value is to be used as bargaining chips in cross-license agreements.

The patent thickets that each company sets up operate like the nuclear missiles held by the U.S. and USSR during the Cold War. Each had enough missiles pointed at the other to create "mutually assured destruction" in the case that one side attacked. Hence, neither side could risk an attack.

It's the same issue with patent thickets. If IBM tries to sue HP for patent infringement, HP would pull out a collection of its own patents and countersue IBM for infringement in some other technology. Even companies that don't particularly want to patent aspects of their business are forced to do so in order to acquire the ammunition necessary for defense against other suits.

The "nuclear bomb" option in patent thickets is a "preliminary injunction." In certain circumstances, a judge might compel a company to stop selling an item that may be infringing on someone else's patent. This can be exceedingly costly. In 1986, Kodak had to completely shut down its instant photography business due to a court-ordered injunction. Eventually Kodak had to pay a billion-dollar judgment for patent infringement.

An injunction to stop production can be a huge threat, but it has no force against companies that don't produce anything. InterTrust, for example, didn't sell any products—all of its income came from licensing patents. Hence, it could threaten to sue other companies for patent infringement without much worry about the threat of countersuits.

EXAMPLE: Managing the Supply of Potatoes

Everyone is familiar with the Organization of Petroleum Exporting Countries (OPEC), the international oil cartel that attempts to influence the price of oil by setting production quota. Normally, coordinating production to push up prices is illegal in the United States, but there are some industries that are exempt from antitrust rules.

A notable example is agricultural producers. The 1922 Capper-Volstead Act specifically exempts farmers from federal antitrust rules. The result has been the creation of a number of "agricultural marketing boards" that attempt to voluntarily regulate the supply of agricultural products.

For example, the United Potato Growers of America, formed in March 2005, has signed up potato farmers that represent over 60% of the potato acreage in the United States. In 2005 it claimed to reduce production of potatoes by 6.8 million sacks of potatoes, each weighing about 100 pounds a piece. According to the *Wall Street Journal* this is equivalent to about 1.3 billion orders of french fries.[2]

25.6 Natural Monopoly

We have seen earlier that the Pareto efficient amount of output in an industry occurs where price equals marginal cost. A monopolist produces where marginal revenue equals marginal cost and thus produces too little output. It would seem that regulating a monopoly to eliminate the inefficiency is pretty easy—all the regulator has to do is to set price equal to marginal

2 Timothy W. Martin, "This Spud's Not for You," *Wall Street Journal*, September 26, 2009.

cost, and profit maximization will do the rest. Unfortunately, this analysis leaves out one important aspect of the problem: it may be that the monopolist would make negative profits at such a price.

An example of this is shown in Figure 25.6. Here the minimum point of the average cost curve is to the right of the demand curve, and the intersection of demand and marginal cost lies underneath the average cost curve. Even though the level of output y_{MC} is efficient, it is not profitable. If a regulator set this level of output, the monopolist would prefer to go out of business.

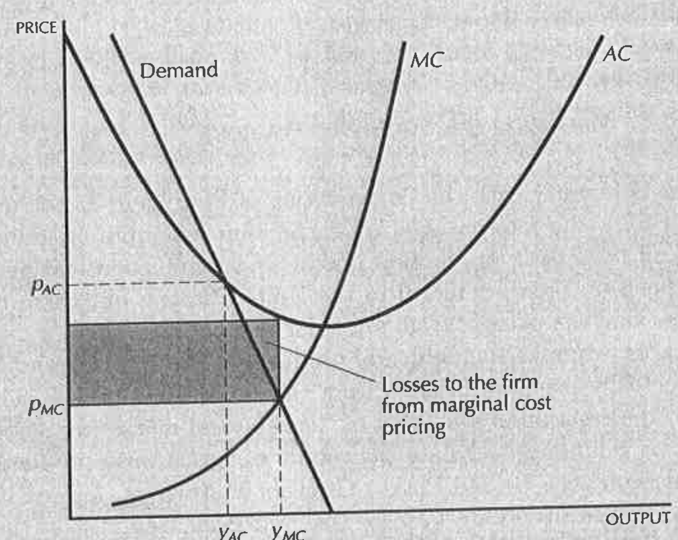

Figure 25.6

A natural monopoly. If a natural monopolist operates where price equals marginal cost, then it will produce an efficient level of output, y_{MC}, but it will be unable to cover its costs. If it is required to produce an output where price equals average cost, y_{AC}, then it will cover its costs, but will produce too little output relative to the efficient amount.

This kind of situation often arises with public utilities. Think of a gas company, for example. Here the technology involves very large fixed costs—creating and maintaining the gas delivery pipes—and a very small marginal cost to providing extra units of gas—once the pipe is laid, it costs very little to pump more gas down the pipe. Similarly, a local telephone company involves very large fixed costs for providing the wires and switching network, while the marginal costs of an extra unit of telephone service is very low. When there are large fixed costs and small marginal costs, you can easily get the kind of situation described in Figure 25.6. Such a situation is referred to as a **natural monopoly**.

If allowing a natural monopolist to set the monopoly price is undesirable due to the Pareto inefficiency, and forcing the natural monopoly to produce at the competitive price is infeasible due to negative profits, what is left? For the most part natural monopolies are regulated or operated by governments. Different countries have adopted different approaches. In some countries the telephone service is provided by the government and in others it is provided by private firms that are regulated by the government. Both of these approaches have their advantages and disadvantages.

For example, let us consider the case of government regulation of a natural monopoly. If the regulated firm is to require no subsidy, it must make nonnegative profits, which means it must operate on or above the average cost curve. If it is to provide service to all who are willing to pay for it, it must also operate on the demand curve. Thus the natural operating position for a regulated firm is a point like (p_{AC}, y_{AC}) in Figure 25.6. Here the firm is selling its product at the average cost of production, so it covers its costs, but it is producing too little output relative to the efficient level of output.

This solution is often adopted as a reasonable pricing policy for a natural monopolist. Government regulators set the prices that the public utility is allowed to charge. Ideally these prices are supposed to be prices that just allow the firm to break even—produce at a point where price equals average costs.

The problem facing the regulators is to determine just what the true costs of the firm are. Usually there is a public utility commission that investigates the costs of the monopoly in an attempt to determine the true average cost and then sets a price that will cover costs. (Of course, one of these costs is the payment that the firm has to make to its shareholders and other creditors in exchange for the money they have loaned to the firm.)

In the United States these regulatory boards operate at the state and local level. Typically electricity, natural gas, and telephone service operate in this way. Other natural monopolies like cable TV are usually regulated at the local level.

The other solution to the problem of natural monopoly is to let the government operate it. The ideal solution here in this case is to operate the service at price equals marginal cost and provide a lump-sum subsidy to keep the firm in operation. This is often the practice for local public transportation systems such as buses and subways. The lump-sum subsidies may not reflect inefficient operation *per se* but rather, simply reflect the large fixed costs associated with such public utilities.

Then again, the subsidies may just represent inefficiency! The problem with government-run monopolies is that it is almost as difficult to measure their costs as it is to measure the costs of regulated public utilities. Government regulatory commissions that oversee the operations of public utilities often subject them to probing hearings to require them to justify cost data whereas an internal government bureaucracy may escape such intense scrutiny. The government bureaucrats who run such government monopolies may turn out to be less accountable to the public than those who run the regulated monopolies.

25.7 What Causes Monopolies?

Given information on costs and demand, when would we predict that an industry would be competitive and when would we predict that it would be monopolized? In general the answer depends on the relationship between the average cost curve and the demand curve. The crucial factor is the size of the **minimum efficient scale (MES)**, the level of output that minimizes average cost, relative to the size of demand.

Consider Figure 25.7 where we have illustrated the average cost curves and the market demand curves for two goods. In the first case there is room in the market for many firms, each charging a price close to p^* and each operating at a relatively small scale. In the second market, only one firm can make positive profits. We would expect that the first market might well operate as a competitive market and that the second would operate as a monopolist.

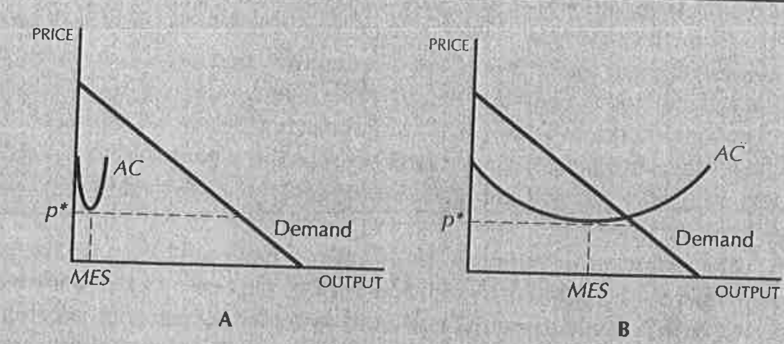

Figure 25.7

Demand relative to minimum efficient scale. (A) If demand is large relative to the minimum efficient scale, a competitive market is likely to result. (B) If it is small, a monopolistic industry structure is possible.

Thus the shape of the average cost curve, which in turn is determined by the underlying technology, is one important aspect that determines whether a market will operate competitively or monopolistically. If the minimum efficient scale of production—the level of output that minimizes average costs—is small relative to the size of the market, we might expect that competitive conditions will prevail.

Note that this is a *relative* statement: what matters is the scale relative to the market size. We can't do too much about the minimum efficient scale—that is determined by the technology. But economic policy can influence the size of the market. If a country chooses nonrestrictive foreign-trade policies, so that domestic firms face foreign competition, then the domestic firms' ability to influence prices will be much less. Conversely, if a country adopts restrictive trade policies, so that the size of the market is limited only to that country, then monopolistic practices are more likely to take hold.

If monopolies arise because the minimum efficient scale is large relative to the size of the market, and it is infeasible to increase the size of the market, then the industry is a candidate for regulation or other sorts of government intervention. Of course such regulation and intervention are costly too. Regulatory boards cost money, and the efforts of the firm to satisfy the regulatory boards can be quite expensive. From society's point of view, the question should be whether the deadweight loss of the monopoly exceeds the costs of regulation.

A second reason why monopoly might occur is that several different firms in an industry might be able to collude and restrict output in order to raise prices and thereby increase their profits. When firms collude in this way and attempt to reduce output and increase price, we say the industry is organized as a **cartel**.

Cartels are illegal. The Antitrust Division of the Justice Department and the Bureau of Competition of the Federal Trade Commission are charged with searching for evidence of noncompetitive behavior on the part of firms. If the government can establish that a group of firms attempted to restrict output or engaged in certain other anticompetitive practices, the firms in question can be forced to pay heavy fines.

On the other hand, an industry may have one dominant firm purely by historical accident. If one firm is first to enter some market, it may have enough of a cost advantage to be able to discourage other firms from entering the industry. Suppose, for example, that there are very large "tooling-up" costs to entering an industry. Then the incumbent—the firm already in the industry—may under certain conditions be able to convince potential entrants that it will cut its prices drastically if they attempt to enter the industry. By preventing entry in this manner, a firm can eventually dominate a market. We will study an example of pricing to prevent entry in Chapter 29.

EXAMPLE: Diamonds Are Forever

The De Beers diamond cartel was formed by Sir Ernest Oppenheimer, a South African mine operator, in 1930. It has since grown into one of the world's most successful cartels. De Beers handles over 80% of the world's yearly production of diamonds and has managed to maintain this near-monopoly for several decades. Over the years, De Beers has developed several mechanism to maintain control of the diamond market.

First, it maintains considerable stocks of diamonds of all types. If a producer attempts to sell outside the cartel, De Beers can quickly flood the market with the same type of diamond, thereby punishing the defector from the cartel. Second, large producers' quotas are based on the *proportion* of total sales. When the market is weak, everyone's production quota is reduced proportionally, thereby automatically increasing scarcity and raising prices.

Third, De Beers is involved at both the mining and wholesaling levels of diamond production. In the wholesale market diamonds are sold to cutters in boxes of assorted diamonds: buyers take a whole box or nothing—they cannot choose individual stones. If the market is weak for a certain size of diamond, De Beers can reduce the number of those diamonds offered in the boxes, thereby making them more scarce.

Finally, De Beers can influence the direction of final demand for diamonds by the $110 million a year it spends on advertising. Again, this advertising can be adjusted to encourage demand for the types and sizes of diamonds that are in relatively scarce supply.[3]

EXAMPLE: Pooling in Auction Markets

Adam Smith once said "People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices." Bidding pools in auctions provide an illustrative example of Smith's observation. In 1988 the Justice Department charged 12 Philadelphia antique dealers with antitrust violations for their participation in this particular kind of "conspiracy against the public."[4]

3 A short description of the diamond market can be found in "The cartel lives to face another threat," *The Economist*, January 10, 1987, 58–60. A more detailed description can be found in Edward J. Epstein, *Cartel* (New York: Putnam, 1978).

4 See Meg Cox, "At Many Auctions, Illegal Bidding Thrives As a Longtime Practice Among Dealers," *Wall Street Journal*, February 19, 1988, which served as the source for this example.

The dealers were accused of participating in "bidding rings," or "pools," at antique furniture auctions. The members of a pool would appoint one member to bid on certain items. If this bidder succeeded in acquiring an item, the participating dealers would then hold a subsequent private auction, called a "knockout," in which the members of the pool bid among themselves for the item. This practice allowed the members of the pool to acquire the items at much lower prices than would have prevailed if they had bid separately; in many cases the prices in the knockout auctions were 50 to 100 percent greater than the prices paid to the original sellers of the goods.

The dealers were surprised by the Justice Department suit; they considered pooling a common business practice in their trade and did not think it was illegal. They thought of the pools as a tradition of cooperation among themselves; being invited to join a pool was considered a "mark of distinction." According to one dealer, "The day I was allowed to go into the pool was a banner day. If you weren't in the pool, you weren't considered much of a dealer." The dealers were so naive that they kept careful records of their payments in the knockout auctions, which were later used by the Justice Department in the suits against the dealers.

The Justice Department argued "if they are joining together to hold down the price [received by the seller] that is illegal." The Justice Department view prevailed over that of the dealers: 11 of the 12 dealers pleaded guilty and settled the matter with fines of $1,000 to $50,000 and probation. The dealer who held out for a jury trial was found guilty and sentenced to 30 days of house arrest and a fine of $30,000.

EXAMPLE: Price Fixing in Computer Memory Markets

DRAM chips are the "dynamic random access memory" chips that go in your computer. They are pretty much an undifferentiated commodity product and the market for DRAMs is (usually) highly competitive. However, there are allegations that several DRAM producers conspired to fix prices and charge computer makers a higher price than would have obtained under purely competitive conditions. Apple Computer, Compaq, Dell, Gateway, HP, and IBM were apparently affected by this conspiracy.

The Department of Justice started investigating these allegations in 2002. In September 2004, Infineon, a German DRAM manufacturer, pleaded guilty to charges of price fixing, and agreed to pay a $160 million fine. This was the third largest criminal fine ever imposed by the Department of Justice's antitrust division.

According to the court documents, Infineon was charged with "Participating in meetings, conversations, and communications with competitors to discuss the prices of DRAM to be sold to certain customers; Agreeing to price levels of DRAM to be sold to certain customers; Exchanging information on sales of DRAM to certain customers, for the purpose of monitoring and enforcing the agreed-upon prices."

Subsequently, four executives at Infineon were sentenced to prison terms and had to pay hefty fines. The antitrust authorities take price fixing very seriously, and the consequences to companies and individuals that engage in such activities can be severe.

Summary

1. When there is only a single firm in an industry, we say that it is a monopoly.

2. A monopolist operates at a point where marginal revenue equals marginal cost. Hence a monopolist charges a price that is a markup on marginal cost, where the size of the markup depends on the elasticity of demand.

3. Since a monopolist charges a price in excess of marginal cost, it will produce an inefficient amount of output. The size of the inefficiency can be measured by the deadweight loss—the net loss of consumers' and the producer's surplus.

4. A natural monopoly occurs when a firm cannot operate at an efficient level of output without losing money. Many public utilities are natural monopolies of this sort and are therefore regulated by the government.

5. Whether an industry is competitive or monopolized depends in part on the nature of technology. If the minimum efficient scale is large relative to demand, then the market is likely to be monopolized. But if the minimum efficient scale is small relative to demand, there is room for many firms in the industry, and there is a hope for a competitive market structure.

REVIEW QUESTIONS

1. The market demand curve for heroin is said to be highly inelastic. Heroin supply is also said to be monopolized by the Mafia, which we assume to be interested in maximizing profits. Are these two statements consistent?

2. The monopolist faces a demand curve given by $D(p) = 100 - 2p$. Its cost function is $c(y) = 2y$. What is its optimal level of output and price?

3. The monopolist faces a demand curve given by $D(p) = 10p^{-3}$. Its cost function is $c(y) = 2y$. What is its optimal level of output and price?

4. If $D(p) = 100/p$ and $c(y) = y^2$, what is the optimal level of output of the monopolist? (Be careful.)

5. A monopolist with constant marginal cost is producing where $|\epsilon| = 3$. The government imposes a quantity tax of \$6 per unit of output. If the demand curve facing the monopolist is linear, how much does the price rise?

6. What is the answer to the above question if the demand curve facing the monopolist has constant elasticity?

7. If the demand curve facing the monopolist has a constant elasticity of 2, then what will be the monopolist's markup on marginal cost?

8. The government is considering subsidizing the marginal costs of the monopolist described in the question above. What level of subsidy should the government choose if it wants the monopolist to produce the socially optimal amount of output?

9. Show mathematically that a monopolist always sets its price above marginal cost.

10. True or false? Imposing a quantity tax on a monopolist will always cause the market price to increase by the amount of the tax.

11. What problems face a regulatory agency attempting to force a monopolist to charge the perfectly competitive price?

12. What kinds of economic and technological conditions are conducive to the formation of monopolies?

APPENDIX

Define the revenue function by $r(y) = p(y)y$. Then the monopolist's profit-maximization problem is

$$\max\ r(y) - c(y).$$

The first-order condition for this problem is simply

$$r'(y) - c'(y) = 0,$$

which implies that marginal revenue should equal marginal cost at the optimal choice of output.

Differentiating the definition of the revenue function gives $r'(y) = p(y)+p'(y)y$, and substituting this into the monopolist's first-order condition yields the alternative form

$$p(y) + p'(y)y = c'(y).$$

The second-order condition for the monopolist's profit-maximization problem is

$$r''(y) - c''(y) \leq 0.$$

This implies that

$$c''(y) \geq r''(y)$$

or that the slope of the marginal cost curve exceeds the slope of the marginal revenue curve.