University of Zurich[UZH]

Leseaufträge «Mikroökonomik I»

Modul 4: Marktformen

Unit 2:

– Monopol

Quellen:

- **Chapter 13 – Monopoly**
  Frank, Robert H, & Cartwright, Edward. (2016). *Microeconomics and Behaviour (2nd European ed.)*. London: McGraw-Hill Education.

CHAPTER

13

MONOPOLY

Virtually every cinema charges different admission prices to moviegoers who belong to different groups. Students pay one price, adults another, senior citizens still another. Some cinemas sell 'ten-packs' of movie tickets at a lower unit price than the tickets they sell at the door. And people who attend showings in the lunch hour sometimes pay much less than those who attend evening showings. None of these practices would be expected under our model of perfect competition, which holds that all buyers pay a single price for a completely standardized product (the so-called *law of one price*).

The same cinema operators who charge different ticket prices to different groups follow quite another practice when it comes to the sale of concession items. Here, the law of one price almost always prevails. Students, adults, senior citizens, international football stars, the clergy, service station attendants and all other patrons pay exactly the same price for their popcorn. The same observation applies to the prices of soft drinks and coffee. These prices, however, are usually much higher than we see for the same items sold in grocery stores and other retail establishments, certainly far greater than any reasonable measure of the marginal cost of providing them.

Both behaviours—charging differential admission prices on the one hand and uniformly high concession prices on the other—are, as we will see, perfectly consistent with what the economic model predicts about the single seller of a good or service.

CHAPTER PREVIEW

In this chapter, our task will be to examine the market structure that least resembles perfect competition—namely, *monopoly*, a market served by a single seller of a product with no close substitutes. We will discuss five factors that lead to this market structure: (1) control over key inputs, (2) economies of scale, (3) patents, (4) network economies, and (5) government licences. We will then see that the monopolist's rule for maximizing

profits in the short run is the same as the one used by perfectly competitive firms. The monopolist will expand output if the gain in revenue exceeds the increase in costs, and will contract output if the loss in revenue is smaller than the reduction in costs.

Our next step will be to examine the efficiency properties of the standard monopoly equilibrium. Unlike the perfectly competitive case, the monopoly equilibrium does not exhaust the potential gains from exchange. In general, the value to society of an additional unit of output will exceed the cost to the monopolist of the resources required to produce it. This finding has often been interpreted to mean that monopoly is less efficient than perfect competition. But this interpretation, we will see, is of only limited practical significance because the conditions that give rise to monopoly are rarely compatible with those required for perfect competition.

Our policy focus in the chapter will be on the question of how the government should treat natural monopolies—markets characterized by downward-sloping long-run average cost curves. We will consider five policy alternatives: (1) state ownership, (2) private ownership with government price regulation, (3) competitive bidding by private firms for the right to be the sole provider of service, (4) vigorous enforcement of antitrust laws designed to prevent monopoly, and finally, (5) a complete *laissez-faire,* or hands-off, policy. Because problems are inherent in each alternative, the best policy generally will be different in different circumstances.

We will finish by examining the monopolist's behaviour when confronted with the options of selling in several separate markets. Here again, the logic of cost–benefit analysis will provide a convenient framework for analysing the firm's decision about whether to alter its current behaviour.

DEFINING MONOPOLY

Monopoly is a market structure in which a single seller of a product with no close substitutes serves the entire market. This definition could hardly appear any simpler, and yet it turns out to be exceedingly difficult to apply in practice.

Consider the example of cinemas with which the chapter began. Is a local cinema a monopoly under our definition? In smaller cities, at least, it is likely to be the only one showing a given film at a given time. Whether it is a monopoly obviously depends on what we mean by a close substitute. If, for example, the cinema is currently showing *Halloween Part 8,* there is likely to be a rich variety of close substitutes for its product. Indeed, literally hundreds of low-grade blood-and-gore films are released each year, and the potential patrons of such films generally do not have to look far if they are dissatisfied with the films available at any particular cinema. But what about a cinema that is in the midst of an exclusive six-month, first-run engagement of the latest *Spiderman* film? For fans of this series, there is really no close substitute. Those who want to see it while the excitement level surrounding its release is still high have only one seller to deal with.

Empirically, one practical measure for deciding whether a firm enjoys significant monopoly power is to examine the cross-price elasticity of demand for its closest substitutes. In one famous antitrust case, the DuPont Corporation was charged with having an effective monopoly on the sale of cellophane. Even though the company sold more than 80 per cent of all cellophane traded, it was able to defend itself against this charge by arguing that the cross-price elasticities between cellophane and its close substitutes—at the time, mainly waxed paper and aluminium foil—were sufficiently high to justify lumping all of these flexible-wrap products into a single market. DuPont sold less than 20 per cent of total industry output under this broader market definition. In a controversial decision, the court deemed that small enough to sustain effective competition.

This is not to say, however, that cross-price elasticity provides a clear, unambiguous measure that distinguishes a product with close substitutes from one without. While there may not be anything quite like the latest *Spiderman* movie, there have always been lots of alternative ways to entertain oneself for two hours. For the person whose heart is set on seeing *Spiderman,* the cinema is a monopolist, but for the person merely out in search of a good movie, the same cinema faces stiff competition. The difference between perfect competition and monopoly often boils down to the question of which of these two types of buyer is more numerous. As in so many other cases in economics, the task of distinguishing between competition and monopoly remains as much an art as a science.

Note carefully that the distinction between monopoly and competition does not lie in any difference between the respective price elasticities of the *market* demand curves for the two cases. On the contrary, the demand curve facing a monopolist is often more elastic than the market demand curve in a perfectly competitive industry. The important distinction between monopoly and competition is that the *demand curve facing the individual competitive firm is horizontal* (irrespective of the price elasticity of the corresponding market demand curve), while the *monopolist's demand curve is simply the downward-sloping demand curve for the entire market.*

FIVE SOURCES OF MONOPOLY

How does a firm come to be the only one that serves its market? Economists discuss five factors, any one or combination of which can enable a firm to become a monopoly. Let us consider these factors in turn.

1. Exclusive Control over Important Inputs The Perrier Corporation of France sells bottled mineral water. It spends millions of euros each year advertising the unique properties of this water, which are the result, it says, of a once-in-eternity confluence of geological factors that created their mineral spring. Most are unable to tell the difference between tap water and Perrier. But others feel differently, and for many of them there is simply no satisfactory substitute for Perrier. Perrier's monopoly position with respect to these buyers is the result of its exclusive control over an input that cannot easily be duplicated.

A similar monopoly position has resulted from the deBeers exclusive control over most of the world's supply of raw diamonds. Synthetic diamonds have now risen in quality to the point where they can occasionally fool even an experienced jeweller. But for many buyers, the preference for a stone that was mined from the earth is not a simple matter of greater hardness and refractive brilliance. They want *real* diamonds, and deBeers is the company that has them.

Exclusive control of key inputs is not a guarantee of permanent monopoly power. The preference for having a real diamond, for example, is based largely on the fact that mined diamonds have historically been genuinely superior to synthetic ones. But assuming that synthetic diamonds eventually do become completely indistinguishable from real ones, there will no longer be any basis for this preference. And as a result, deBeers' control over the supply of mined diamonds will cease to confer monopoly power. New ways are constantly being devised of producing existing products, and the exclusive input that generates today's monopoly is likely to become obsolete tomorrow.

2. Economies of Scale As discussed in the previous chapter, when the long-run average cost curve (given fixed input prices) is downward sloping, the least costly way to serve the market is to concentrate production in the hands of a single firm. In Figure 13.1, for example, note that a single firm can produce an industry output of Q^* at an average cost of LAC_{Q^*}, while with two firms sharing the same market, average cost rises to $LAC_{Q^*/2}$. A market that is most cheaply served by a single firm is called a *natural monopoly.* A frequently cited example is the provision of electricity to households.

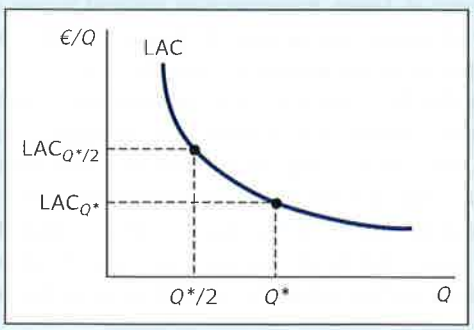


FIGURE 13.1
Natural Monopoly
When the LAC curve is declining throughout, it is always cheaper for a single firm to serve the entire industry.

Recall from Chapter 12 that it is possible for the LAC curve to be downward sloping even in the absence of economies of scale. This can happen, for example, if the price of an important input falls significantly when industry output expands (a *pecuniary economy,* in the language of Chapter 12). Note carefully, however, that this case is *not* one that gives rise to natural monopoly. Input prices here depend on the level of industry output, not on the output of any one firm. Pecuniary economies will apply with equal force whether one or many firms serve the market.

Strictly speaking, then, it is the degree of returns to scale, not the slope of the LAC curve, that determines whether we have a natural monopoly. With fixed input prices, of course, there is always a one-to-one relationship between returns to scale and the slope of the LAC curve (see Chapter 11).

3. Patents Most countries protect inventions through some sort of patent system. A patent typically confers the right to exclusive benefit from all exchanges involving the invention to which it applies. There are costs as well as benefits to patents. On the cost side, the monopoly it creates usually leads, as we will see, to higher prices for consumers. On the benefit side, the patent makes possible a great many inventions that would not otherwise occur. Although some inventions are serendipitous, most are the result of long effort and expense in sophisticated research laboratories. If a firm were unable to sell its product for a sufficiently high price to recoup these outlays, it would have no economic reason to undertake research and development in the first place. Without a patent, competition would force price down to marginal cost, and the pace of innovation would be slowed dramatically.

A case in point is the microprocessor that powers personal computers and a growing array of other products. The fixed investment required to produce the latest Intel chip is several billion euros. Once the chip has been designed and the manufacturing facility built, however, the marginal cost of producing each chip is only a few cents. The protection from competition afforded by a patent is what makes it possible for the firm to recover its costs of innovation. In the EU, the life of a patent is 20 years, a compromise figure that is too long for many inventions, too short for many others. In particular, there is a persuasive argument that the patent life should be extended in the prescription drug industry, where the testing and approval process often consumes all but a few years of the current patent period.[1]

ECONOMIC NATURALIST 13.1

What is a just reward for designing the iPhone or finding a treatment for cancer?

In 2011 Apple launched a series of lawsuits against competitors for patent infringement. Companies, including Samsung and Motorola, were accused of using patented technology on smartphones and tablets. It was not long before Samsung counter-sued—arguing that Apple had infringed its patents. Expensive legal battles in the US, Europe and the Far East ensued.

To illustrate the complexity, consider just one product—Samsung's Galaxy Tab 10.1 tablet computer. A German court ruled in favour of Apple and banned the tablet from sale. A Dutch court originally ruled in favour of Apple, but the verdict was later overruled. In the UK, the court ruled that Apple should put a disclaimer on its website stating that Samsung did not breech Apple's patent; the judge said that the Galaxy Tab was not 'cool' enough to be confused with the iPad. Very confusing! Needless to say, appeals followed.

Most attention, however, was inevitably focused on the much broader case in the US. Here, the court ruled largely in favour of Apple. Samsung was ordered to pay over $1 billion in damages to Apple for infringing various patents such as the 'Bounce-Back Effect' and 'Tap to Zoom'.

Apple would argue that patent protection gave it the incentive to be innovative. Its lawyer Harold McElhinny told the jury that 'In . . . three months, Samsung was able to copy and incorporate the result of Apple's four-year investment in hard work and ingenuity—without taking any of the risks'. The key thing here is risk—many projects fail and so a firm must be confident that on the successful projects it can recoup not only the costs from the successful project but also all the unsuccessful projects. Without patent protection we may never have seen the iPhone or iPad.

Others, however, would argue that Apple has done pretty well out of the iPhone and iPad. Patent protection is now merely stifling competition, and keeping lawyers happy. Was patent

[1]Henry Grabowski, *Drug Regulation and Innovation,* Washington, DC: American Enterprise Institute, 1976.

protection key in Apple's decision to invest in the iPhone? Probably not. But that is only because Apple operates in markets where there is significant brand loyalty. When the latest iPad or iPhone is released customers queue for hours to get their hands on one. These customers will clearly buy the iPhone and iPad even if Samsung offers a similar, cheaper product. Apple profits, therefore, from being a market leader, irrespective of patent laws.

Things are different in markets with no brand loyalty. In the prescription drug industry, for example, patent protection is crucial. In this market customers can easily substitute to cheaper brands. There would, therefore, be no incentive for Pfizer, Novartis or AstraZeneca to discover a treatment for cancer if they knew that their rivals could undercut them and run away with all the profits. ■

4. Network Economies On the demand side of many markets, a product becomes more valuable as greater numbers of consumers use it.[2] A vivid early illustration was the VHS technology's defeat of the competing Beta format in home video recorders. The attraction of VHS over the initial versions of Beta was that it permitted longer recording times. Beta later corrected this deficiency, and on most important technical dimensions became widely regarded by experts as superior to VHS. Yet the initial sales advantage of VHS proved insuperable. Once the fraction of consumers owning VHS passed a critical threshold, the reasons for choosing it became compelling—variety and availability of tape rentals, access to repair facilities, the capability to exchange tapes with friends, and so on.

In extreme cases, such *network economies* function like economies of scale as a source of natural monopoly. Microsoft's Windows operating system, for example, achieved its dominant market position on the strength of powerful network economies. Because Microsoft's initial sales advantage gave software developers a strong incentive to write for the Windows format, the inventory of available software in the Windows format became vastly larger than for any competing operating system. And although general-purpose software such as word processors and spreadsheets continue to be available for multiple operating systems, specialized professional software usually appear first in the Windows format and often only in that format. This software gap gave people a good reason for choosing Windows, even if, as in the case of many Apple Macintosh users, they believe a competing system is otherwise superior.

The end result is that around 85 per cent of the world's personal computers now run Microsoft's Windows operating system. If that's not a pure monopoly, it comes awfully close. Microsoft's monopoly position is, however, increasingly under threat. Between them, Android and Apple's iOS are the dominant operating systems in mobile phones and tablets. This has turned the tables on Microsoft, with software increasingly being written for operating systems other than Windows. A monopoly caused by network economies can, therefore, be transient. Just as VHS has seen its day, maybe Windows has too.

5. Government Licences or Franchises In many markets, the law prevents anyone but a government-licensed firm from doing business. At service areas on motorways, for example, not just any fast-food restaurant is free to set up operations. The motoring authorities negotiate with several companies, choose one, and then grant it an exclusive licence to serve a particular area. The authorities' purpose in restricting access in the first place is that there is simply not room for more than one establishment in these locations. In such cases, the government licence as a source of monopoly is really a scale economy acting in another form. But government licences are also required in a variety of other markets, such as the one for taxis, where scale economies do not seem to be an important factor.

Government licences are sometimes accompanied by strict regulations that spell out what the licensee can and cannot do. Where the government gives a chain restaurant an exclusive licence, for example, the restaurant will often be required to charge prices no more than, say, 10 per cent higher than it charges in its unregulated outlets. In other cases, the government simply charges an extremely high fee for the licence, virtually forcing the licensee to charge premium prices. This is the

[2]See, for example, Joseph Farrell and Garth Saloner, 'Standardization, Compatibility, and Innovation', *Rand Journal of Economics*, 16, 1985: 70–83; and M. L. Katz and Carl Shapiro, 'Systems Competition and Network Effects', *Journal of Economic Perspectives*, Spring 1994: 93–115.

practice of some airport authorities, who essentially auction their terminal counter space to the highest bidders. Your annoyance at having to pay £5 for a coffee in Heathrow or Gatwick Airports is thus more properly focused on Heathrow Airport Holdings or GIP, which owns Gatwick.

By far the most important of the five factors for explaining monopolies that endure is economies of scale. Production processes change over time, which makes exclusive control over important inputs only a transitory source of monopoly. Patents too are inherently transitory. Network economies, once firmly entrenched, can be as persistent a source of natural monopoly as economies of scale. Strictly speaking, network economies work through the demand side of the market by affecting what buyers are willing to pay for a product. But they may be equivalently conceptualized on the supply side as yet another feature of product quality. The more people who own the product, the higher its effective quality level. It may thus be said of a product that benefits from network economies that any given quality level can be produced at lower cost as sales volume increases. Viewed in this way, network economies are just another form of economies of scale in production, and that is how we shall view them in the discussion that follows. Government licences can persist for extended periods, but many of these licences are themselves merely an implicit recognition of scale economies that would lead to monopoly in any event.

With this brief overview of the causes of monopoly in mind, let us turn now to the question of what the consequences of monopoly are. In order to do this, we will proceed in much the same fashion as we did in our study of the competitive firm. That is, we will examine the firm's output decision and ask whether it leads to a situation in which all possible gains from exchange are exhausted. The answer is generally no. But in formulating a government policy to improve on the results of unregulated monopoly, we will see that it is critical to understand the original source of monopoly.

THE PROFIT-MAXIMIZING MONOPOLIST

As in the competitive case, we assume that the monopolist's goal is to maximize economic profit. And again as before, in the short run this means to choose the level of output for which the difference between total revenue and short-run total cost is greatest. The case for this motive is less compelling than in the case of perfect competition. After all, the monopolist's survival is less under siege than the competitor's, and so the evolutionary argument for profit maximization applies with less force in the monopoly case. Nonetheless, we will explore just what behaviours follow from the monopolist's goal of profit maximization.

The Monopolist's Total Revenue Curve

The key difference between the monopolist and the perfect competitor is the way in which total, and hence marginal, revenue varies with output. Recall from Chapter 12 that the demand curve facing the perfect competitor is simply a horizontal line at the short-run equilibrium market price—call it P^*. The competitive firm is a price taker, typically because its own output is too small to have any discernible influence on the market price. Under these circumstances, the perfectly competitive firm's total revenue curve is a ray with slope P^*, as shown in Figure 13.2.

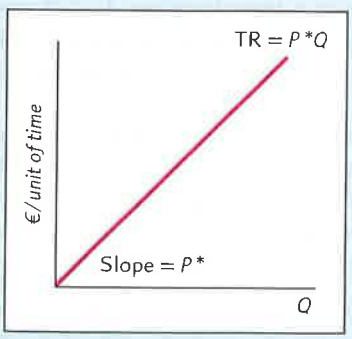


FIGURE 13.2
The Total Revenue Curve for a Perfect Competitor
Price for the perfect competitor remains at the short-run equilibrium level P^* irrespective of the firm's output. Its total revenue is thus the product of P^* and the quantity it sells: $TR = P^*Q$.

Now consider a monopolist with the downward-sloping demand curve $P = 80 - (\frac{1}{5})Q$ pictured in the top panel in Figure 13.3. For this firm as well, total revenue is the product of price and quantity. At point A on its demand curve, for example, it sells 100 units of output per week at a price of €60/unit, giving a total revenue of €6,000/wk. At B, it sells 200 units at a price of €40, so its total revenue at B will be €8,000/wk, and so on. The difference between the monopolist and the competitor is that for the monopolist to sell a larger amount of output, it must cut its price—not only for the marginal unit but for all preceding units as well. As we saw in Chapter 6, the effect of a downward-sloping demand curve is that total revenue is no longer proportional to output sold. As in the competitive case, the monopolist's total revenue curve (middle panel in Figure 13.3) passes through the origin, because in each case selling no output generates no revenue. But as price falls, total revenue for the monopolist does not rise linearly with output. Instead, it reaches a maximum value at the quantity corresponding to the midpoint of the demand curve (B in the top panel), after which it again begins to fall. The corresponding values of the price elasticity of demand are shown in the bottom panel in Figure 13.3. Note that total revenue reaches its maximum value when the price elasticity of demand is unity.

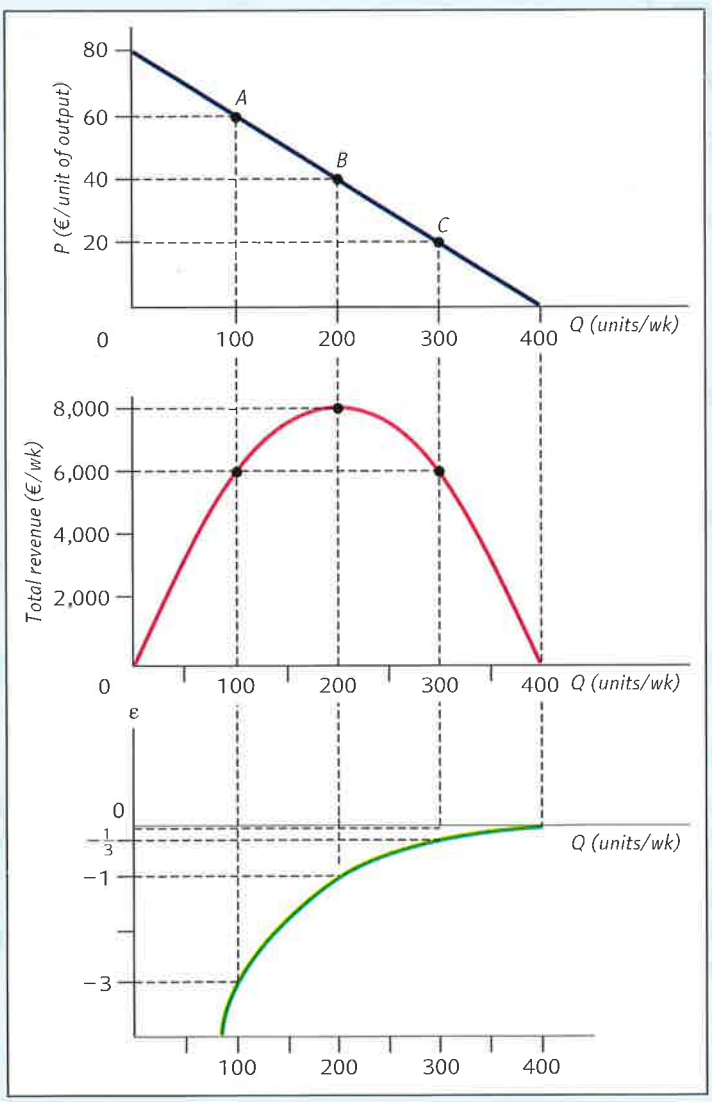


FIGURE 13.3
Demand, Total Revenue and Elasticity
For the monopolist to increase sales, it is necessary to cut price (top panel). Total revenue rises with quantity, reaches a maximum value, and then declines (middle panel). The quantity level for which the price elasticity of demand is unity corresponds to the midpoint of the demand curve, and at that value total revenue is maximized.

EXERCISE 13.1 **Sketch the total revenue curve for a monopolist whose demand curve is given by $P = 100 - 2Q$.**

The top panel in Figure 13.4 portrays the short-run total cost curve and total revenue curve for a monopolist facing the demand curve shown in Figure 13.3. Economic profit, plotted in the bottom panel, is positive in the interval from $Q = 45$ to $Q = 305$, and is negative elsewhere. The maximum profit point occurs at $Q^* = 175$ units/wk, which lies to the left of the output level for which total revenue is at a maximum ($Q = 200$).

Notice in Figure 13.4 that the vertical distance between the short-run total cost and total revenue curves is greatest when the two curves are parallel (when $Q = 175$). Suppose this were not the case. For example, suppose that at the maximum-profit point the total cost curve were steeper than the total revenue curve. It would then be possible to earn higher profits by producing less output, because costs would go down by more than the corresponding reduction in total revenue. Conversely, if the total cost curve were less steep than the total revenue curve, the monopolist could earn higher profits by expanding output, because total revenue would go up by more than total cost.

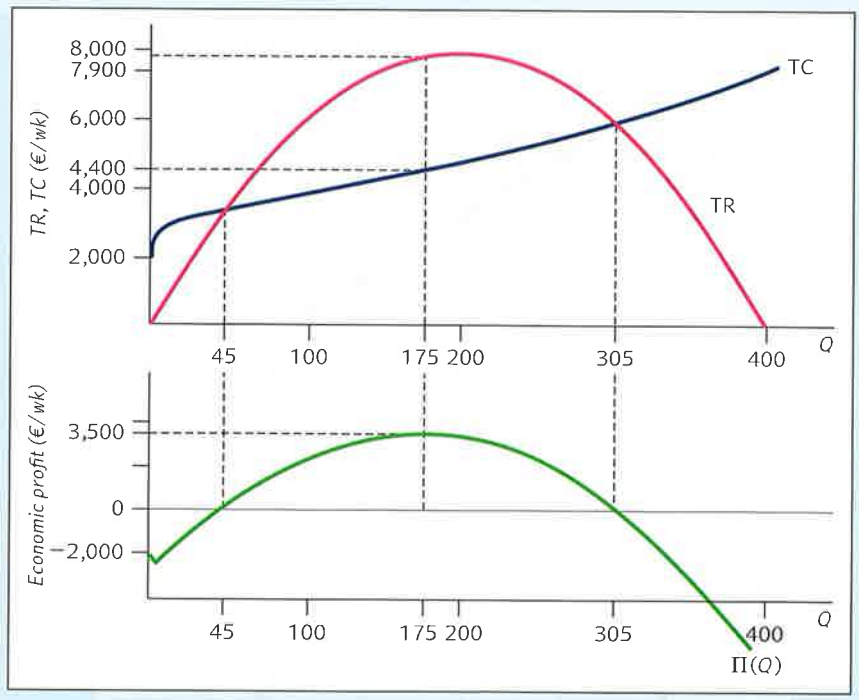


FIGURE 13.4
Total Cost, Revenue and Profit Curves for a Monopolist
Economic profit ($\Pi(Q)$ in the bottom panel) is the vertical distance between total revenue and total cost (TR and TC in the top panel). Note that the maximum-profit point, $Q^* = 175$, lies to the left of the output level at which TR is at a maximum ($Q = 200$).

Marginal Revenue

The slope of the total cost curve at any level of output is by definition equal to marginal cost at that output level. By the same token, the slope of the total revenue curve is the definition of *marginal revenue.*[3] As in the case of the perfectly competitive firm, we can think of marginal revenue as the change in total revenue when the sale of output changes by 1 unit. More precisely, suppose ΔTR_Q is the change in total revenue that occurs in response to a small change in output, ΔQ. Marginal revenue, denoted MR_Q, is then given by

$$MR_Q = \frac{\Delta TR_Q}{\Delta Q} \tag{13.1}$$

[3]In calculus terms, marginal revenue is defined as the derivative dTR/dQ.

Using this definition, a profit-maximizing monopolist in the short run will choose that level of output Q^* for which

$$MC_{Q^*} = MR_{Q^*} \tag{13.2}$$

provided marginal revenue intersects marginal cost from above. Equation 13.2 defines the **optimality condition for a monopolist**. The monopolist wants to sell all units for which marginal revenue exceeds marginal cost, so marginal revenue should lie above marginal cost for output below the optimum. (For some cost structures, marginal cost may decline initially and then increase, leading to two intersections of marginal cost and marginal revenue.)

optimality condition for a monopolist a monopolist maximizes profit by choosing the level of output where marginal revenue equals marginal cost.

Recall that the analogous condition for the perfectly competitive firm is to choose the output level for which price and marginal cost are equal. Recalling that marginal revenue and price are exactly the same for the competitive firm (when such a firm expands output by 1 unit, its total revenue goes up by P), we see that the profit-maximizing condition for the perfectly competitive firm is simply a special case of Equation 13.2.

In the case of the monopoly firm, marginal revenue will always be less than price.[4] To see why, consider the demand curve pictured in Figure 13.5, and suppose that the monopolist wishes to increase output from $Q_0 = 100$ to $Q_0 + \Delta Q = 150$ units/wk. His total revenue from selling 100 units/wk is (€60/unit) (100 units/wk) = €6,000/wk. To sell an additional $\Delta Q = 50$ units/wk, he must cut his price to €60 − ΔP = €50/unit, which means his new total revenue will be (€50/unit) (150 units/wk), which is equal to €7,500/wk. To calculate marginal revenue, we simply subtract the original total revenue, €6,000/wk, from the new total revenue, and divide by the change in output, $\Delta Q = 50$ units/wk. This yields $MR_{Q_{0=100}}$ = (€7,500/wk − €6,000/wk)/(50 units/wk) = €30/unit, which is clearly less than the original price of €60/unit.

Another useful way of thinking about marginal revenue is to view it as the gain in revenue from new sales minus the loss in revenue from selling the previous output level at the new, lower price. In Figure 13.5, the area of rectangle B (€2,500/wk) represents the gain in revenue from the additional sales at the lower price. The area of rectangle A (€1,000/wk) represents the loss in revenue from selling the original 100 units/wk at €50/unit instead of €60. Marginal revenue is the difference between the gain in revenue from additional sales and the loss in revenue from sales at a lower price, divided by the change in quantity. This yields (€2,500/wk − €1,000/wk)/(50 units/wk), which is again equal to €30/unit.

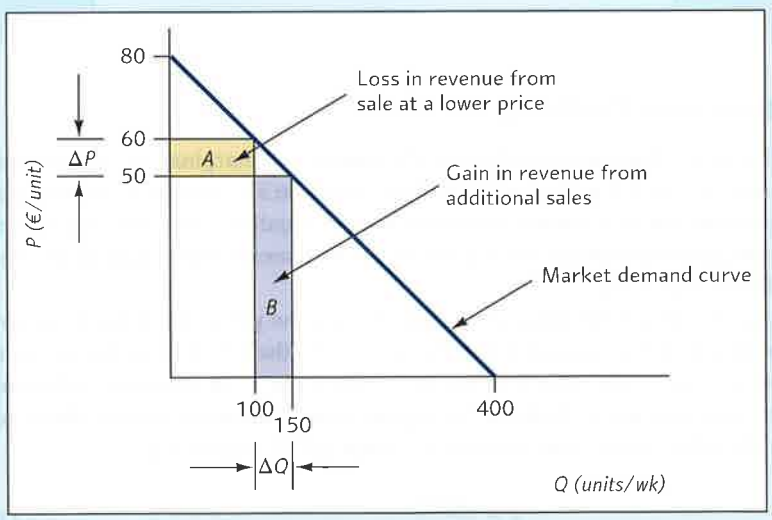


FIGURE 13.5
Changes in Total Revenue Resulting from a Price Cut
The area of rectangle A (€1,000/wk) is the loss in revenue from selling the previous output level at a lower price. The area of rectangle B (€2,500/wk) is the gain in revenue from selling the additional output at the new, lower price. Marginal revenue is the difference between these two areas (€2,500 − €1,000 = €1,500/wk) divided by the change in output (50 units/wk). Here, MR equals €30/unit, which is less than the new price of €50/unit.

[4]There is actually one exception to this claim, namely, the case of the perfectly discriminating monopolist, discussion of which follows.

To explore how marginal revenue varies as we move along a straight-line demand curve, consider the demand curve pictured in Figure 13.6, and suppose that the monopolist wishes to increase output from Q_0 to $Q_0 + \Delta Q$ units. His total revenue from selling Q_0 units is P_0Q_0. To sell an additional ΔQ units, he must cut his price to $P_0 - \Delta P$, which means his new total revenue will be $(P_0 - \Delta P)(Q_0 + \Delta Q)$, which is equal to $P_0Q_0 + P_0\Delta Q - \Delta PQ_0 - \Delta P\Delta Q$. To calculate marginal revenue, simply subtract the original total revenue, P_0Q_0, from the new total revenue, and divide by the change in output, ΔQ. This leaves $MR_{Q_0} = P_0 - (\Delta P/\Delta Q)Q_0 - \Delta P$, which is clearly less than P_0. As ΔP approaches zero, the expression for marginal revenue thus approaches[5]

$$MR_{Q_0} = P_0 - \frac{\Delta P}{\Delta Q} Q_0 \tag{13.3}$$

Equation 13.3 makes intuitive sense if we think of ΔQ as being a 1-unit change in output; P_0 would then be the gain in revenue from the sale of that extra unit, and $(\Delta P/\Delta Q)Q_0 = \Delta PQ_0$ would be the loss in revenue from the sale of the existing units at the lower price. We see again in Equation 13.3 that marginal revenue is less than price for all positive levels of output.

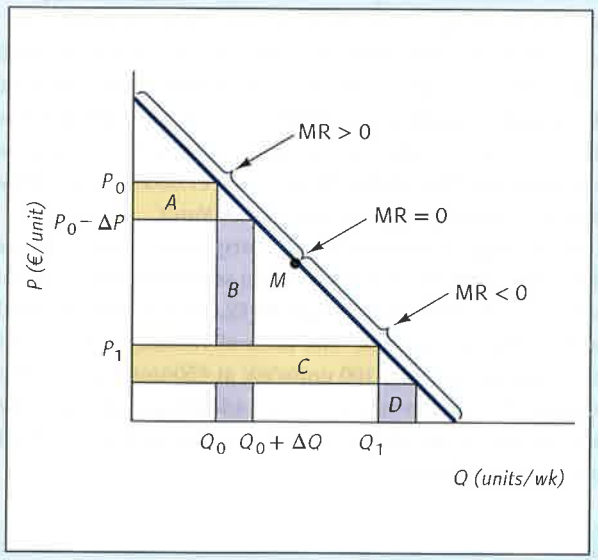


FIGURE 13.6
Marginal Revenue and Position on the Demand Curve
When Q is to the left of the midpoint (M) of a straight-line demand curve (for example, $Q = Q_0$), the gain from added sales (area B) outweighs the loss from a lower price for existing sales (area A). When Q is to the right of the midpoint (for example, $Q = Q_1$), the gain from added sales (area D) is smaller than the loss from a lower price for existing sales (area C). At the midpoint of the demand curve, the gain and the loss are equal, which means marginal revenue is zero.

Marginal Revenue and Elasticity

The fact that area B is larger than area A in Figure 13.6 means that marginal revenue is positive at Q_0. Once output moves past the midpoint (M in Figure 13.6) on a straight-line demand curve, however, the marginal revenue of a further expansion will be negative. Thus, the area of rectangle C is larger than the area of rectangle D in Figure 13.6, which means that marginal revenue at the output level Q_1 is less than zero.

Recall from Chapter 6 that total demand is highest when the price elasticity of demand is equal to 1. This corresponds to the midpoint M in Figure 13.6. To the left of M we know that marginal revenue is positive and to the right it is negative. This hints at the important relationship between marginal revenue and price elasticity. To explore this relationship in more detail, recall from Chapter 6 that the price elasticity of demand at a point (Q, P) is given by

$$\varepsilon = \frac{\Delta Q}{\Delta P} \frac{P}{Q} \tag{13.4}$$

[5]Note that when ΔP shrinks toward zero, the corresponding ΔQ does so as well. Because ΔP and ΔQ are both positive here, the ratio $\Delta P/\Delta Q$ is simply the negative of the slope of the demand curve.

In Equation 13.4, the terms ΔQ and ΔP have opposite signs, because the demand curve is downward sloping. By contrast, recall that the ΔQ and ΔP terms in Equation 13.3, which also represent changes in P and Q as we move along the demand curve, are both positive. Suppose we redefine ΔQ and ΔP from Equation 13.4 so that both of these terms are positive. That equation then becomes

$$|\varepsilon| = \frac{\Delta Q}{\Delta P}\frac{P}{Q} \tag{13.5}$$

The purpose of making both ΔQ and ΔP positive is to be able to relate Equation 13.5 back to Equation 13.3. If we now solve Equation 13.5 for $\Delta P/\Delta Q$ $|\varepsilon|$ and substitute into Equation 13.3, we get

$$\text{MR}_Q = P\left(1 - \frac{1}{|\varepsilon|}\right) \tag{13.6}$$

Equation 13.6 tells us that the less elastic demand is with respect to price, the more price will exceed marginal revenue.[6] It also tells us that in the limiting case of infinite price elasticity, marginal revenue and price are exactly the same. (Recall from Chapter 12 that price and marginal revenue are the same for the competitive firm, which faces a horizontal, or infinitely elastic, demand curve.)

Graphing Marginal Revenue

Equation 13.6 provides a convenient way to plot the marginal revenue values that correspond to different points along a demand curve. To illustrate, consider the straight-line demand curve in Figure 13.7, which intersects the vertical axis at a price value of $P = 80$. The elasticity of demand is infinite at that point, which means that $\text{MR}_0 = 80(1/|\varepsilon|) = 80$. Although marginal revenue will generally be less than price for a monopolist, the two are exactly the same when quantity is zero. The reason is that at zero output there are no existing sales for a price cut to affect.

Now suppose we move, say, one-quarter of the way down the demand curve to point A, (100, 60). At that point, $|\varepsilon| = 3$. Thus we have $\text{MR}_{100} = (60)\left(1 - \frac{1}{3}\right) = 40$.

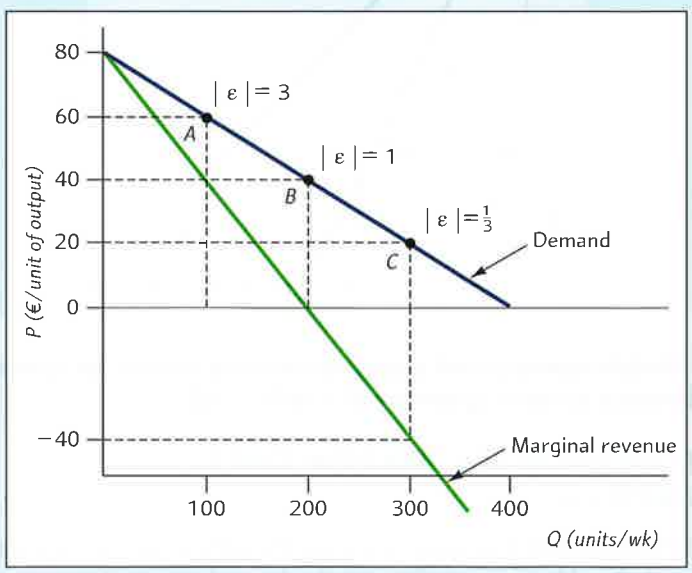


FIGURE 13.7
The Demand Curve and Corresponding Marginal Revenue Curve
For the case of a straight-line demand curve, the corresponding marginal revenue curve is also a straight line. It has the same vertical intercept as the demand curve, and its horizontal intercept is half that of the demand curve.

[6]Equation 12.6 can be derived using calculus as follows:

$$\text{MR} = \frac{d\text{TR}}{dQ} = \frac{d(PQ)}{dQ} = P + Q\frac{dP}{dQ} = P\left(1 + \frac{QdP}{PdQ}\right) = P\left(1 + \frac{1}{\varepsilon}\right) = P\left(1 - \frac{1}{|\varepsilon|}\right)$$

Halfway down the demand curve, at point B, (200, 40), $|\varepsilon| = 1$, which gives us $MR_{200} = (40)\left(1 - \frac{1}{1}\right) = 0$. This confirms our earlier finding (Chapter 6) that total revenue is at a maximum at the midpoint of a straight-line demand curve, where elasticity is unity.

Finally, consider point C, (300, 20), which is three-fourths of the way down the demand curve. Here $|\varepsilon| = \frac{1}{3}$, so we have $MR_{300} = (20)[1 - (1/\frac{1}{3})] = (20)(-2) = -40$. Thus, at $Q = 300$, the effect of selling an extra unit of output is to reduce total revenue by €40/wk.

Filling in additional points in the same fashion, we quickly see that the marginal revenue curve associated with a straight-line demand curve is itself a straight line, one whose slope is twice that of the demand curve. The marginal revenue curve cuts the horizontal axis just below the midpoint of the demand curve, and for all quantities larger than that, marginal revenue is negative. Note that all points to the right of the midpoint of the demand curve have price elasticity values less than 1 in absolute value. The fact that marginal revenue is negative in this region thus fits our observation from Chapter 6 that a cut in price will reduce total revenue whenever demand is inelastic with respect to price.

EXAMPLE 13.1 Find the marginal revenue curve that corresponds to the demand curve $P = 12 - 3Q$.

The marginal revenue curve will have the same intercept as and twice the slope of the demand curve, which gives us $MR = 12 - 6Q$, as plotted in Figure 13.8.

The general formula for a linear demand curve is $P = a - bQ$, where a and b are positive numbers. The corresponding marginal revenue curve will be $MR = a - 2bQ$.[7] ◆

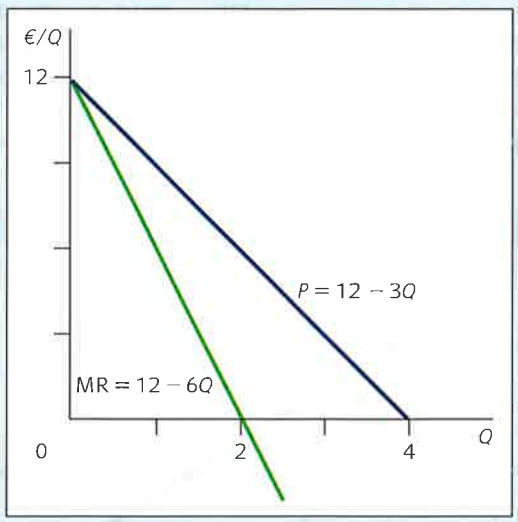


FIGURE 13.8
A Specific Linear Demand Curve and the Corresponding Marginal Revenue Curve
The marginal revenue curve has the same vertical intercept and twice the slope of the corresponding linear demand curve.

EXERCISE 13.2 Sketch demand and marginal revenue curves for a monopolist whose market demand curve is given by $P = 100 - 2Q$.

Graphical Interpretation of the Short-Run Profit Maximization Condition

Recall (see Equation 13.2) that a profit maximizing monopolist will produce where marginal revenue equals marginal cost. Consider a monopolist with the demand, marginal revenue and short-run cost curves pictured in Figure 13.9. The profit-maximizing level of output for this firm is Q^*,

[7]Note that total revenue for the demand curve $P = a - bQ$ is given by $TR = aQ - bQ^2$. The corresponding marginal revenue curve is

$$MR = \frac{dTR}{dQ} = a - 2bQ$$

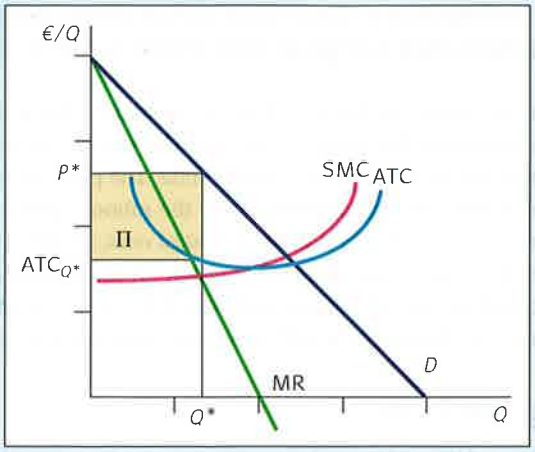


FIGURE 13.9
The Profit-Maximizing Price and Quantity for a Monopolist
Maximum profit occurs at the output level Q^*, where the gain in revenue from expanding output (or loss in revenue from contracting output), MR, is exactly equal to the cost of expanding output (or the savings from contracting output), SMC. At Q^*, the firm charges P^* and earns an economic profit of Π.

the one for which the marginal revenue and marginal cost curves intersect. At that quantity level, the monopolist can charge a price of P^*, and by so doing will earn an economic profit equal to the shaded rectangle labelled Π.

EXAMPLE 13.2 A monopolist faces a demand curve of $P = 100 - 2Q$ and a short-run total cost curve of $TC = 640 + 20Q$. What is the profit-maximizing price? How much will the monopolist sell, and how much economic profit will it earn at that price?

The marginal revenue curve for this demand curve is $MR = 100 - 4Q$. Marginal cost is the slope of the total cost curve, which is constant at 20 in this example. Setting MR = MC, we have $100 - 4Q = 20$, which yields the profit-maximizing quantity, $Q^* = 20$. Plugging $Q^* = 20$ back into the demand curve, we get the profit-maximizing price, $P^* = 60$. This solution is shown graphically in Figure 13.10, which also displays the average total cost curve for the monopolist. Note that at Q^* the ATC is 52, which means the monopolist earns an economic profit of $60 - 52 = 8$ on each unit sold. With $Q^* = 20$, that makes for a total economic profit of 160. ◆

Note in Figure 13.10 that the monopolist's fixed cost was irrelevant to the determination of the profit-maximizing output level and price. This makes sense intuitively, because fixed cost has no bearing on the gains and losses that occur when output changes.

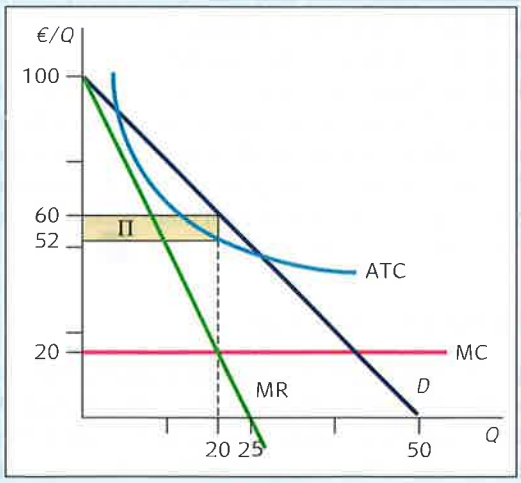


FIGURE 13.10
The Profit-Maximizing Price and Quantity for Specific Cost and Demand Functions

EXERCISE 13.3 **How would the profit-maximizing price and quantity change in Example 13.2 if the monopolist's total cost curve were instead given by TC = 640 + 40*Q*? The associated marginal cost curve is MC = 40.**

If a monopolist's goal is to maximize profits, it follows directly that she will never produce an output level on the inelastic portion of her demand curve. If she were to increase her price at such an output level, the effect would be to increase total revenue. The price increase would also reduce the quantity demanded, which, in turn, would reduce the monopolist's total cost. Since economic profit is the difference between total revenue and total cost, profit would necessarily increase in response to a price increase from an initial position on the inelastic portion of the demand curve. The profit-maximizing level of output must therefore lie on the elastic portion of the demand curve, where further price increases would cause both revenue and costs to go down.

The Profit-Maximizing Mark-Up

The profit-maximization condition MR = MC can be combined with Equation 13.6, which says MR = $P[1 - (1/|\varepsilon|)]$, to derive the profit-maximizing mark-up for the monopolist:

$$\frac{P - \text{MC}}{P} = \frac{1}{|\varepsilon|} \tag{13.7}$$

which is the difference between price and marginal cost, expressed as a fraction of the profit-maximizing price. For example, if the price elasticity of demand facing a monopolist were equal to −2, the profit-maximizing mark-up would be 1/2, which implies that the profit-maximizing price is twice marginal cost. Equation 13.7 tells us that the profit-maximizing mark-up grows smaller as demand grows more elastic. In the limiting case of infinitely elastic demand, the profit-maximizing mark-up is zero (which implies P = MC), the same as in the perfectly competitive case.

ECONOMIC NATURALIST 13.2

Why is wine more expensive in a restaurant than a supermarket?

Once diners are seated in a restaurant they typically have only one choice if they want to drink wine with their meal. They have to pay whatever the restaurant charges. This gives the restaurant an effective monopoly on the supply of wine. The inevitable consequence is a price well above marginal cost. By contrast, someone looking for a bottle of wine to drink at home can buy wine from any number of places (except for Sweden). Competition between supermarkets and local stores pushes price down towards marginal cost.

It is no surprise, therefore, that wine costs considerably more in a restaurant than supermarket. It is not uncommon, for instance, that the same bottle of wine would cost €30 in a restaurant and only €10 in the supermarket next door. Note that this equates to a profit-maximizing mark-up of 2/3. Which would suggest the price elasticity of demand for wine is −1.5.

One interesting aspect of the restaurant–supermarket comparison is that customers are keenly aware of it. The typical diner *knows* that wine, and all other drinks, are priced well above marginal cost. Some (see Chapter 8) may consider this unfair and simply not buy wine because they view it as 'over-priced'. In other cases the profit-maximizing mark-up is less transparent. Legoland Hotel in Billund, for example, has a monopoly for Lego-themed hotel rooms with direct access to Legoland. Given the unique nature of this hotel there is no direct comparator. Customers, therefore, do not know just how over-priced the rooms are. ■

The Monopolist's Shutdown Condition

In the case of the perfectly competitive firm, we saw that it paid to shut down in the short run whenever the price fell below the minimum value of average variable cost (AVC). The analogous condition for the monopolist is that there exists no quantity for which the demand curve lies above the average variable cost curve. The monopolist whose demand, marginal revenue, SMC and AVC curves are shown in Figure 13.11, for example, has no positive level of output for which

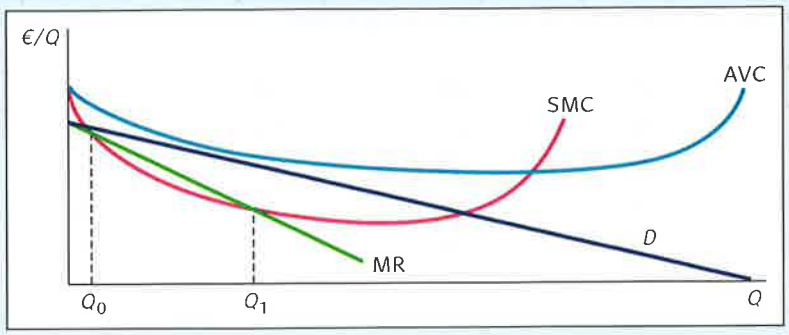


FIGURE 13.11
A Monopolist Who Should Shut Down in the Short Run
Whenever average revenue (the price value on the demand curve) is lower than average variable cost for every level of output, the monopolist does best to cease production in the short run.

price exceeds AVC, and so the monopolist does best by ceasing production in the short run. He will then sustain a short-run economic loss equal to his fixed costs, but he would do even worse at any positive level of output.

Another way of stating the shutdown condition for a monopolist is to say that he should cease production whenever average revenue is less than average variable cost at every level of output. Average revenue is simply another name for price—the value of P along the monopolist's demand curve.[8]

Figure 13.11 also illustrates the important point that MR = MC is a necessary, but not sufficient, condition for maximum profit. Note in the figure that marginal revenue is equal to marginal cost at the output level Q_0. Why isn't this the maximum-profit point? Recall that in the case of the perfectly competitive firm, the maximum-profit condition called for price to equal marginal cost on a rising portion of the marginal cost curve, above the minimum point on the AVC curve. A somewhat different condition applies in the case of the monopolist. In Figure 13.11, note that at Q_0 the MR curve intersects the MC curve from below.[9] This means not only that Q_0 is not the maximum-profit point, but that it actually corresponds to a *lower* profit level than any of the other output levels nearby. For example, consider an output level just less than Q_0. At any such output level the gains from contracting output (MC) will exceed the losses (MR), so the firm does better to contract from Q_0. Now consider an output level just slightly larger than Q_0. For such an output level, the gains from expanding (MR) exceed the costs (MC), so the firm does better to expand. Thus, when the firm is at Q_0, it can earn higher profits by either contracting *or* expanding. Q_0 is called a *local minimum* profit point.[10]

Note also in Figure 13.11 that the MR curve intersects the MC curve a second time at the output level Q_1. This time the intersection occurs from above, and you can easily show as an exercise that Q_1 yields higher profits than any of the other output levels close by. (The argument runs exactly parallel to the one in the preceding paragraph.) We refer to points like Q_1 as *local maximum* profit points. But although Q_1 yields more profit than any nearby output level, the firm fails to cover its average variable cost at the level of output, and so does better simply to produce nothing at all. The point Q^* we saw earlier in Figure 13.9 is both a local maximum profit point and a *global maximum* profit point, the latter designation indicating that no other output level, including zero, yields higher profit. For a monopolist, a global maximum profit point might occur either on the rising or on the falling portion of the MC curve. But it must be at a point where the MR curve intersects the MC curve from above.

[8]More formally, note that average revenue = $TR/Q = PQ/Q = P$.
[9]To 'intersect from below at Q_0' means that as Q approaches Q_0 from the left, MR lies below MC and then crosses MC when $Q = Q_0$.
[10]The second-order condition for maximum profit is given by

$$\frac{d(\text{MR} - \text{MC})}{dQ} = \frac{d\text{MR}}{dQ} - \frac{d\text{MC}}{dQ} < 0$$

which says simply that the slope of the marginal revenue curve must be less than the slope of the marginal cost curve.

EXERCISE 13.4 **Find the optimal price and quantity for the monopolist described by the information in the following table.**

Q	P	MR	SMC	AVC
0	100	100	150	150
15	86	71	71	107
25	75	50	41	84
34	66	33	33	72
50	50	0	63	63

To recapitulate briefly, we have seen that the monopolist behaves like a perfectly competitive firm in the sense that each chooses an output level by weighing the benefits of expanding (or contracting) output against the corresponding costs. For both the perfect competitor and the monopolist, marginal cost is the relevant measure of the cost of expanding output. Fixed costs are irrelevant for short-run output decisions in both cases. For both the monopolist and the perfect competitor, the benefits of expanding output are measured by their respective values of marginal revenue. For the competitor, marginal revenue and price are one and the same. For the monopolist, by contrast, marginal revenue is less than price. The competitor maximizes profit by expanding output until marginal cost equals price. The monopolist maximizes profit by expanding output until marginal cost equals marginal revenue, and thus chooses a lower output level than if he had used the competitor's criterion. Both the monopolist and the perfect competitor do best to shut down in the short run if price is less than average variable cost for all possible levels of output.

A MONOPOLIST HAS NO SUPPLY CURVE

As we saw in Chapter 12, the competitive firm has a well-defined supply curve. It takes market price as given and responds by choosing the output level for which marginal cost and price are equal. At the industry level, a shifting demand curve will trace out a well-defined industry supply curve, which is the horizontal summation of the individual firm supply curves.

There is no similar supply curve for the monopolist. The reason is that the monopolist is not a price taker, which means that there is no unique correspondence between price and marginal revenue when the market demand curve shifts. Thus, a given marginal revenue value for one demand curve can correspond to one price, while the same value of marginal revenue for a second demand curve corresponds to a different price. As a result, it is possible to observe the monopolist producing Q_1^* and selling at P^* in one period, and then selling Q_2^* at P^* in another period.

To illustrate, consider a monopolist with a demand curve of $P = 100 - Q$ and with the same cost curves as in Example 13.2, in particular with MC = 20. The marginal revenue curve for this monopolist is given by MR = $100 - 2Q$, and equating MR to MC yields a profit-maximizing output level of $Q^* = 40$. The corresponding profit-maximizing price is $P^* = 60$. Note that this is the same as the profit-maximizing price we saw for the monopolist in Example 13.2, even though the demand curve here ($P = 100 - Q$) lies to the right of the earlier one ($P = 12 - 3Q$). In Example 13.2 we saw that the profit-maximizing output level was $Q^* = 20$. Hence, a price of 60 can be consistent with output of both 40 and 20 depending on the demand curve.

When the monopolist's demand curve shifts, the price elasticity of demand at a given price generally will also shift. But these shifts need not occur in the same direction. When demand shifts rightward, for example, elasticity at a given price may either increase or decrease, and the same is true when demand shifts leftward. The result is that there can be no unique correspondence between the price a monopolist charges and the amount she chooses to produce. And hence we say that the monopolist has no supply curve. Rather, she has a *supply rule,* which is to equate marginal revenue and marginal cost.

ECONOMIC NATURALIST 13.3

What is the optimal price for sweets at Legoland in Windsor?

The natural geography of Legoland means that everyone exiting from the park is funnelled through a relatively narrow stretch of path. And Legoland had enough business acumen to realize that a sweet shop at the narrowest point was a good idea. Children cannot but help notice the shop. That leaves a typical parent with the difficult choice between over-priced sweets or a trip home with a disappointed child (see also Economic Naturalist 9.8).

Put yourself in the position of the manager of this shop deciding what prices to charge. You have a monopoly because there are no other sweet shops nearby. The only thing you need to take into account, therefore, is the willingness of parents to pay for sweets. If you price too low you miss out on profit. If you price too high parents may decide a disappointed child is the best option. So, what you need to do is estimate the demand curve and find the optimal point to price on that demand curve, given your costs.

If the demand curve shifts then you may want to reassess the prices you charge. Everything, though, is driven by demand. This is why it is inappropriate to think of a supply curve relating price and quantity. This is not to say, however, that supply is irrelevant. After all, the fact that the shop is a monopoly tells us a lot about the supply side of this market. ■

ADJUSTMENTS IN THE LONG RUN

In the long run, the monopolist is of course free to adjust all inputs, just as the competitive firm is. What is the optimal quantity in the long run for a monopolist with a given technology? The best the monopolist can do is to produce the quantity for which long-run marginal cost is equal to marginal revenue. In Figure 13.12, that will mean choosing a capital stock that gives rise to the short-run average and marginal cost curves labelled ATC* and SMC*. For that level of capital stock, the short-run marginal cost curve passes through the intersection of the long-run marginal cost and marginal revenue curves. Q^* will be the profit-maximizing quantity in the long run, and it will sell at a price of P^*. For the conditions pictured in Figure 13.12, the long-run economic profit level, Π, will be positive, and is indicated by the area of the shaded rectangle.

As we saw in Chapter 12, economic profits tend to vanish in the long run in perfectly competitive industries. This tendency will sometimes be present for monopoly. To the extent that the factors that gave rise to the firm's monopoly position come under attack in the long run, there will be downward pressure on its profits. For example, competing firms may develop substitutes for important inputs that were previously under the control of the monopolist. Or in the case of patented products, competitors may develop close substitutes that do not infringe on existing patents, which are in any event only temporary.

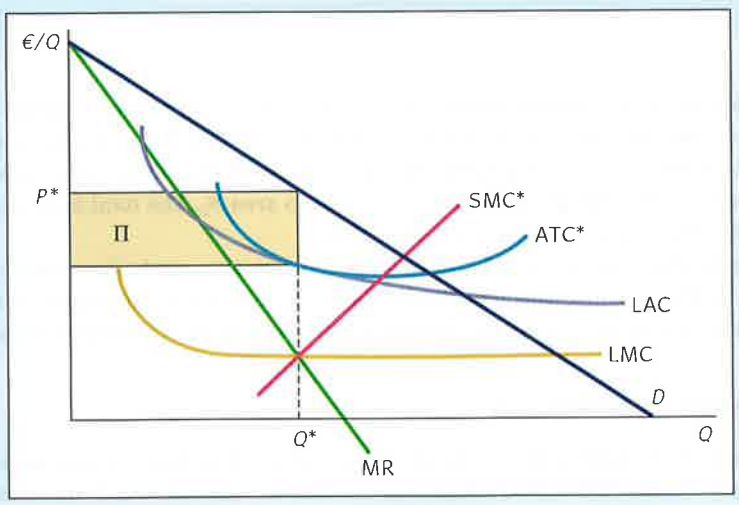


FIGURE 13.12
Long-Run Equilibrium for a Profit-Maximizing Monopolist
The profit-maximizing quantity in the long run is Q^*, the output level for which LMC = MR. The profit-maximizing price in the long run is P^*. The optimal capital stock in the long run gives rise to the short-run marginal cost curve SMC*, which passes through the intersection of LMC and MR.

But in other cases there may be a tendency for monopoly profits to persist. The firm shown in Figure 13.12, for example, has a declining long-run average cost curve, which means that it may enjoy a persistent cost advantage over potential rivals. In such natural monopolies, economic profits may be highly stable over time. And the same, of course, may be true for a firm whose monopoly comes from having a government licence. Persistent economic profits are indeed one of the major policy concerns about monopoly, as we will discuss later in the chapter.

THE EFFICIENCY LOSS FROM MONOPOLY

Recall from Chapter 12 the claim that perfect competition led to an efficient allocation of resources. This claim was based on the observation that in both short-run and long-run competitive equilibrium, there are no possibilities for additional gains from exchange. The value to buyers of the last unit of output is exactly the same as the market value of the resources required to produce it. This is encapsulated in the condition that price equals marginal cost, $P = \text{MC}$.

How does the long-run equilibrium under monopoly measure up by the same criteria? Not very well, it turns out. To illustrate, consider a monopolist with the short-run marginal cost and demand curve shown in Figure 13.13. The profit-maximizing quantity for this monopolist is Q^*, which he will sell at a price of P^*. Note that at Q^*, the value of an additional unit of output to buyers is P^*, which is greater than the cost of producing an additional unit, MC. This means that the single-price monopolist does not exhaust all possible gains from exchange.

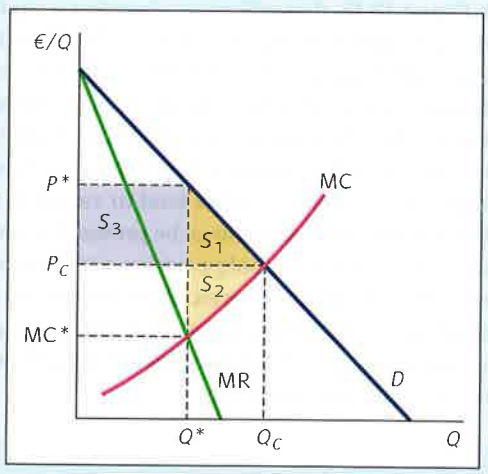


FIGURE 13.13
The Welfare Loss from a Single-Price Monopoly
A monopolist will produce Q^* and sell at P^*. A competitive industry operating under the same cost conditions would produce Q_C and sell at P_C. In comparison with the perfectly competitive outcome, monopoly results in a loss of consumer surplus equal to the area of S_1 and a loss of producer surplus equal to area S_2.

The loss in efficiency can be measured by comparing consumer and producer surplus at the monopoly and efficient outcome. We can see in Figure 13.13 that the efficient outcome would be output Q_C at price P_C. Relative to the monopoly outcome this would result in extra consumer surplus, equal to area S_1, and extra producer surplus, equal to area S_2. The total loss, equal to areas S_1 and S_2, is called the *deadweight loss from monopoly*.

Note that monopoly also results in a redistribution from consumer surplus to producer surplus equal to area S_3 in Figure 13.13. This motivates a fairness objection to monopoly in that a monopolist, by charging a high price, extracts consumer surplus in order to increase profits.

Dynamic Efficiency of Monopoly

In the proceeding analysis, it made sense to speak of the welfare loss from having monopoly rather than competition because the cost structure was one that is compatible with the existence of perfect competition. But with that kind of cost structure only legal barriers could prevent the

emergence of competition. The existence of economic profits should lure competitors into the industry until price and quantity are driven to the efficient level.

But suppose the reason for having a monopoly is that the firm enjoys patent protection for its product. Can we say that the welfare loss from having a monopoly is equal to the deadweight loss measured in Figure 13.13? Before answering, we must first ask, 'What is the alternative to the current situation?' If it is a society without patent protection, we may well have never got the product in the first place, so it hardly makes sense to complain that, compared with perfect competition, monopoly produces a welfare loss. True enough, the patent-protected single-price monopoly does not exhaust all possible gains from trade. But, with the patent-protected monopoly, we do get some consumer and producer surplus, whereas we might have got nothing at all without the patent protection.

A closely related issue is the dynamic use of monopoly profits. Innovation of new products and production techniques is almost always a costly exercise, and economic profits are a key source of funds for the relevant research and development. In the case of a patent protection research and development can be undertaken with the promise of future profits for any breakthrough. This option, however, is only open to firms with access to sufficient funds. A more realistic alternative for most firms is that past profits can be used to fund current research and development. But that means that low profits can delay innovation. And, again, it is not clear that a society which exploits all *current* gains from exchange but fails to invest in *future* product development is efficient. Short-term gain may result in a long-term welfare loss due to a lack of new products or cost-saving production techniques.

Two examples illustrate the point. Consider first the mobile phone market in the EU and the progression from 3G to 4G. The mobile market in most countries is considered competitive, with prices being kept relatively low. This is great news for consumers in the short run. Low prices and profits have, however, delayed investment in the new technology and infrastructure required for 4G and beyond. That is not good news for consumers. As a second example, consider competition in the car market. The car market is also considered relatively competitive within the EU. Again, this is good for consumers in the short run. Low profits, however, threaten development of the hybrid and electric cars that are needed if global emissions are to be reduced in the future. As a case in point, Volkswagon's reaction to the emissions scandal of 2015 was to make clear that spending would have to be cut and that investment in technology and car development was likely to be cut most.

These examples do not imply that monopoly is more efficient than perfect competition. Indeed, one of the most enduring topics of conversation among economic conspiracy buffs is the notion that monopolists deprive consumers of a spectrum of enormously valuable technological innovations. Who has not heard, for example, of how the light bulb manufacturers have conspired to prevent revolutionary new designs for long-lasting light bulbs from reaching the market? The proceeding illustrations should caution, however, against the uncritical acceptance of the idea that monopoly results in an efficiency loss. And as the following example makes clear, monopolists are unlikely to be eager to suppress innovation.

EXAMPLE 13.3 Suppose the current light bulb design lasts 1,000 hours. Now the light bulb monopolist discovers how to make a bulb that lasts 10,000 hours for the same per-bulb cost of production. Will the monopolist introduce the new bulb?

Suppose we measure the quantity produced by the monopolist not as light bulbs per se, but as the number of bulb-hours of lighting services. Thus, if the cost of producing the current design is, say, €1/bulb-hr, then the cost of the new design is only €0.10/bulb-hr. In Figure 13.14, D represents the market demand curve for lighting and MR the associated marginal revenue curve.

Note that the profit-maximizing price and quantity for the current design, whose marginal cost is €1/bulb-hr, are P_1 and Q_1, respectively. For the new design, whose marginal cost is €0.10/bulb-hr, the profit-maximizing price and quantity are P_2 and Q_2. The monopolist's profit under the current design is the area of the rectangle $ABCE$. For the new design, the corresponding

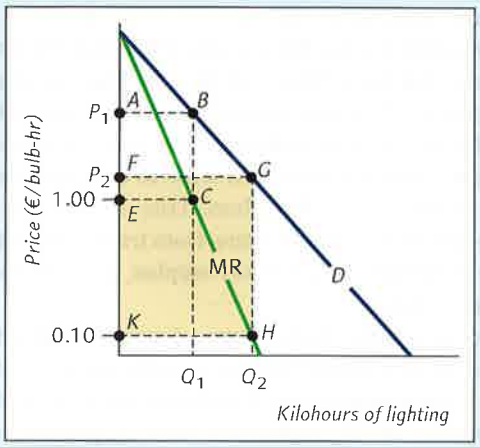


FIGURE 13.14
Does Monopoly Suppress Innovation?
The cost of producing the new, efficient light bulb, at €0.10/bulb-hr, is only one-tenth the cost of producing the current design, €1/bulb-hr. Because the monopolist's profits with the efficient design (area of *FGHK*) exceed its profits with the current design (area of *ABCE*), it will offer the new design.

profit value is the area of the rectangle *FGHK*. And because the monopolist's profit is higher under the new design, it has every incentive to make that design available. Indeed, as some of you may recall, the availability of just such an efficient new light bulb was announced several years ago. ◆

Two other reasons for caution in judging the inefficiency of monopoly are provided by economies of scale and price discrimination. We look at each in turn in the remainder of this chapter.

PUBLIC POLICY TOWARD NATURAL MONOPOLY

The preceding discussion made clear that the relevant question is not whether monopoly is efficient in comparison with some unattainable theoretical ideal, but how it compares with the alternatives we actually confront. This question is nowhere more important than in the case of natural monopoly.

Consider the natural monopoly depicted in Figure 13.15. Here, a profit-maximizing monopolist would produce at Q^*, which is less than the efficient output of Q^{**}. We can still, therefore, talk of a deadweight loss from monopoly (equal to area *S* in Figure 13.15). Note, however, that output

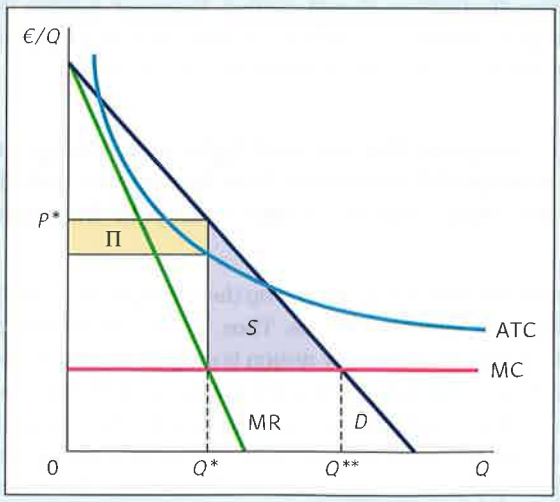


FIGURE 13.15
A Natural Monopoly
The two main objections to single-price natural monopoly are that it earns economic profit (Π) and that it results in the loss of consumer surplus (*S*).

Q^{**} is only possible if there is a monopoly firm. Moreover, any firm that produces output Q^{**} will make an economic loss. Competition, therefore, is not the answer.

Policymakers may respond in a variety of ways to the inefficiency of a profit-maximizing natural monopolist. The five options considered below account for the most important alternatives.

1. State Ownership and Management

Efficiency requires that price be equal to marginal cost. The difficulty this creates is that, for natural monopoly, marginal cost is below average total cost. Because private firms are not able to charge prices less than average cost and remain in business in the long run, the single-price firm has no alternative but to charge more than marginal cost. An option for getting around this particular difficulty is to have the state take over the industry. The attractive feature of this option is that the government is not bound, the way a private firm is, to earn at least a normal profit. It would thus be able to set a price equal to marginal cost, and absorb the resulting economic losses out of general tax revenues.

Traditionally, many natural monopolies have been state owned. Utility companies, for example, that provide water, electricity or gas are often state owned. As are telecom companies, postal services, railway and road providers and television services. In each case state ownership can be motivated by the large fixed costs that would make private provision inefficient.

But there are also unattractive features of state ownership. Foremost among them is the fact that it often seems to weaken incentives for cost-conscious, efficient management. As the late Harvard University economist Harvey Leibenstein emphasized, an organization's costs depend not just on its technology, but also on the vigour with which it pursues efficiency. In Leibenstein's phrase, an organization that does not act energetically to curb costs is said to exhibit **X-inefficiency.**[11]

X-inefficiency a condition in which a firm fails to obtain maximum output from a given combination of inputs.

X-inefficiency is by no means the exclusive province of government. In widely varying degrees, it is found in private firms as well. Leibenstein argued that the extent to which X-inefficiency is a problem will depend on economic incentives, which suggests a theoretical reason for believing that it is likely to be more widespread in government. When a private firm cuts a euro from its costs, its profit goes up by a euro. By contrast, when the person in charge of a government agency cuts a euro from her agency's budget, the effect is merely to shrink her fiefdom.

Several noted scholars have argued that the goal of most bureaucrats is to maximize their operating budgets.[12] This is not to deny that bureaucrats are for the most part sincere, dedicated public servants. But it is perhaps only human nature for a bureaucrat to think that her particular agency has the most important mission in government, and to lobby accordingly on its behalf.

Persuaded by these arguments, the trend in Europe has been towards the privatization of state-owned natural monopolies. The telecom sector, for instance, has largely passed from public to private hands. British Telecom was privatized in 1984, Deutsche Telecom in 1995 and France Telecom in 2004. Electricity and gas markets were next, with state-owned monopolies such as British Gas, Electricité de France and Gaz de France passing to at least partial private ownership. The European Commission has been instrumental in the liberalization of both telecom and energy markets, requiring the opening up of national markets to competition.

The liberalization of the telecom sector is typically heralded as a success. Prices were reduced, efficiency increased, and companies like British Telecom, Deutsche Telecom and France Telecom are competing successfully on a global scale. This success stems partly from a reduction in X-inefficiency caused by privatization. It also stems, however, from an interesting property of natural monopolies—they are often temporary. As technology improves, cost curves change, and natural monopolies can stop being natural monopolies. This is arguably what has

[11]Harvey Leibenstein, 'Allocative Efficiency vs. X-Efficiency', *American Economic Review*, June 1966: 392–415.

[12]See, for example, William Niskanen, *Bureaucracy and Representative Government*, Chicago: Aldine-Atherton, 1971; and Gordon Tullock, *The Politics of Bureaucracy*, Washington, DC: Public Affairs Press, 1965. But for a contrasting view, see Albert Breton and Ronald Wintrobe, *The Logic of Bureaucratic Conduct, Cambridge:* Cambridge University Press, 1982.

happened in the telecom sector, particularly with the advent of mobile phones. Competition is now strong within the telecom sector, making it difficult to gauge how much of the improvement in efficiency is due to privatization or the change in market structure.

The evidence on liberalization within industries that are unambiguously still natural monopolies is more mixed. Prices in the energy sector remain high. And there is no convincing evidence that efficiency is higher in the UK and the Netherlands, where markets are more competitive, than France or Germany, where the state maintains a larger presence. This has led to concerns over further liberalization—as the following Economic Naturalist illustrates.

ECONOMIC NATURALIST 13.4

Why are British railways so bad compared to European railways?

In 1994 British Rail was privatized. This followed an EU directive requiring the *decoupling* of track and infrastructure management from the provision of passenger services. British Rail was split into a company called Railtrack, that would have ownership of things like track, signalling and stations, and 25 passenger train operating units. The basic idea was that the train operating units compete to operate services on a rail system maintained and operated by Railtrack.

Anyone who travels by British Rail cannot but notice that things have improved since privatization. Indeed, some trains now run on time. Privatization has not, however, been without its problems. Following financial difficulties and a fatal train crash, Railtrack was renationalized. And most acknowledge that breaking up the industry into so many companies has pushed up the cost of running the railway.

Having observed the UK's troubles, the German and French governments have shown reluctance to go along with the European Commission's plans for liberalizing rail services. France, for instance, created a new company RFF with ownership of track, and an agency DCF to manage access. The old state monopoly SNCF would maintain the track and run its own services. The latest plans, however, are to bring everything back under SNCF with enough internal separation to satisfy the EU directive on decoupling. This is how things are already done in Germany.

To put things in context, consider Figure 13.16 with a stylized comparison of British Rail and Deutsche Bahn. Suppose that demand is equal to Q^*. This is not to say that demand for trains is perfectly price inelastic—it is not. It more recognizes that rail users and the government typically share the cost of providing rail services. We will suppose the government fixes user prices so that demand is Q^*. British Rail has average total cost curve ATC_{BR} and Deutsche Bahn has curve ATC_{DB}. Clearly costs are a lot higher with British Rail. This is because of X-inefficiency. It means that providing Q^* costs ATC^*_{DB} rather than ATC^*_{DB} resulting in extra spending equal to area *ABCD*. This spending goes on higher salaries, extra staff, bureaucracy, the manager's new office, etc.

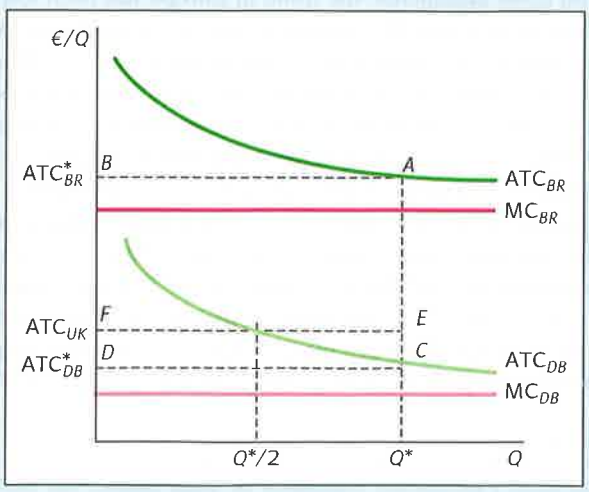


FIGURE 13.16
Costs on European Railways Compared
Costs are significantly higher with the old British Rail (BR) than with Deutsche Bahn (DB). This is due to X-inefficiency. With privatization, costs in the UK become similar to those in Germany. But, the UK market is fragmented resulting in a higher average cost than Germany.

After privatization X-inefficiency is removed, meaning the average total cost curve in the UK is the same as that of Deutsche Bahn. Fragmentation of the industry, however, means that average costs will be relatively high because firms cannot benefit from economies of scale. Figure 13.16 depicts what happens as the industry is split into two firms, each providing $Q^*/2$. Average costs become ATC^*_{UK}. These are substantially lower than with British Rail but still higher than in Germany. The extra cost is now equal to area *CDEF*.

This example emphasizes that the best scenario with a natural monopoly is state ownership and efficient marginal cost pricing. If that is what Deutsche Bahn and SNCF are achieving, then why mess around with it? Privatization may have helped in the UK but it still leaves a 'second-best' outcome. ■

If state-owned companies can avoid X-inefficiency then state-ownership is best.[13] Recent experience arguably shows that X-inefficiency can be avoided in state-owned companies. This helps explain why the tide appears to be turning against the privatization of natural monopolies. Indeed, in the UK, privatization is fast becoming a word that politicians dare not use, which is an abrupt change from 10 or 20 years ago. But, there are several alternatives to state ownership that will still need to be considered.

2. State Regulation of Private Monopolies

One such alternative is to put ownership in private hands, while providing guidelines or regulations that limit pricing discretion. Most of the formerly state-owned monopolies that have been privatized within Europe have been regulated to some extent. In the UK, for instance, Ofwat, Ofgem and Ofcom regulate, respectively, water, energy and telecommunication companies. National regulators cooperate on a European level through associations such as the European Council of European Energy Regulators and Agency for the Cooperation of Energy Regulators.

The simplest, and most common, form of regulation is a cap on prices. There are two basic options open to the regulator. The first is to set price equal to marginal cost. Note, however, that this would result in the firm making an economic loss and so the firm would have to be subsidized by the government. The second option is to set price equal to average cost. The firm is, thus, constrained to earn zero economic profit. We do not get price equal to marginal cost, but we get the next best thing.

If the price (and subsidy) are set at the right level by the regulator then the firm has an incentive to reduce X-inefficiency. Otherwise, it would make an economic loss. The basic difficulty, however, is knowing what price is appropriate. The regulator simply does not know what the marginal cost and average cost curves of the monopolist look like and so is not in a position to know what price would equate marginal or average cost with demand.

If regulators set the price too low the firm will have an incentive to reduce the quality of its service and eventually to go out of business. By contrast, if regulators set the price too high the firm will earn an economic profit. Neither of these outcomes is attractive, but regulators have traditionally decided that the problems caused by a low price are far more serious than those caused by an excessive one.

And clearly the firm has an incentive to overstate its costs if this will result in it being able to set a higher price. For instance, it can increase the salaries of senior managers, buy gold-plated water coolers, finance corporate golf events and pass off such costs as essential to the firm. This is obviously not a recipe for eliminating X-inefficiency.

Regulatory pitfalls have not prevented governments in virtually every part of the world from continuing to intervene in the price and output decisions of important natural monopolies like electric and gas utilities. Whether these interventions do more good than harm, in purely economic terms, remains an unsettled question. But they clearly seem to serve an important psychological function on behalf of a public that feels understandably uncomfortable about not having a buffer between itself and the sole supplier of a critical good or service.

[13]See Elliott D. Sclar, *You Don't Always Get What You Pay For*, Ithaca, NY: Cornell University Press, 2000.

Refuse collection in the UK is routinely franchised out to private companies

3. Exclusive Contracting for Natural Monopoly

The basic problem with regulation is that the firm knows a lot more than the regulator. In particular, the firm knows its cost function, the regulator does not. This constrains the regulator's ability to control the firm. Is there not a way to force the firm to reveal its cost function?

In the title of a widely quoted article, UCLA economist Harold Demsetz asked the disarmingly simple question, 'Why Regulate Monopoly?'[14] His point was that even though cost conditions may dictate that a market be served by a single supplier, there can still be vigorous competition to see who gets to *be* that supplier. In Demsetz's proposal, the government would specify in detail the service it wanted provided—fire protection, garbage collection, postal delivery, whatever—and then call for private companies to submit bids to supply the service. And to the low bidder would then go the contract. The advantage of this contracting out approach is that firms have an incentive to bid the true value of operating in the market. There are not the distorted incentives that we see with regulation.

This scheme has been widely used in the US and the UK for many years. The most high profile example in the UK is rail franchising. Under this system train operating companies compete on a rolling basis for the right to operate passenger services in a particular area or along particular routes. A similar system exists for the running of prisons. Many local services are also contracted out to the private sector—refuse collection, catering in hospitals and schools, etc. Competition for contracts is typically high. This, one would hope, will keep costs down and reduce X-inefficiency.[15]

Private contracting, though, may not be an attractive option in many cases. It will always be tempting, corruption aside, to go with the lowest bid. Many services, however, are complex enough that setting out in detail what is expected of the private provider is impossible. The quality of railway service, for example, depends on punctuality, number of free seats, quality of on-board coffee, etc. The cheapest provider may not provide the best on-board coffee. So, judging who will provide the best value for money is complicated. A further complication is the potential for the provider to go bankrupt mid-contract because it underestimated the cost of providing the service. These problems are all too familiar to those who have followed the rail franchising experience of the UK.

4. Vigorous Enforcement of Antitrust Laws

A major element in the policy arsenal for dealing with monopoly are antitrust laws. Article 102 of the Treaty on the Functioning of the European Union prohibits firms holding a dominant position on a determined market to abuse that position, for example by charging unfair prices, limiting production or refusing to innovate. Article 101 limits agreements between companies that would restrict competition.

In the case of industries with increasing returns to scale, however, the cost of production will be much higher if we are served by many firms rather than by only a few. The most vigorous supporters of the antitrust laws insist that the laws will not impede the formation of natural monopolies. But as we will see in Chapter 14, they may substantially postpone the time when economies of scale are fully realized.

[14]Harold Demsetz, 'Why Regulate Monopoly?' *Journal of Law and Economics*, April 1968: 55–65.
[15]For an extended survey of studies comparing private costs and public costs, see E. S. Savas, *Privatizing the Public Sector*, Chatham, NJ: Chatham House Publishers, 1982.

One response to this difficulty would be to apply the antitrust laws to prevent only those mergers where significant cost savings would not be realized. The European Commission is not in a good position, however, to distinguish one type of case from another. The result is that antitrust policy impedes all consolidations, even those that would lead to substantial reductions in cost.

Is the EU powerless against Microsoft?

ECONOMIC NATURALIST 13.5

We have already mentioned that over 80 per cent of the world's personal computers run Microsoft's Windows operating system. Microsoft's monopoly position clearly gives it power to manipulate the market towards its own ends. The job of competition authorities is to stop it abusing that power.

The easiest way to do that is to stop such a dominant market position from being established in the first place. And the EU competition authorities do regularly stop mergers to prevent this happening. The Microsoft example illustrates, however, how difficult it can be to stop a firm obtaining a monopoly. Microsoft's initial success came from giving customers what they wanted—and that is the supposed benefit of competition. Its continued success reflects the network economies that undoubtedly exist in the market. These network economies create a natural monopoly.

The competition authorities are left, therefore, with the task of making sure Microsoft does not abuse the power it has. Since the turn of the century, Microsoft has been fined well over a billion euros by the EU for anti-competitive behaviour. It has been ordered to make its operating system available to open source software developers. And, it was ordered to unbundle various software from its operating system.[16] These measures check the power of Microsoft, but arguably not much.

The greater threat to Microsoft's dominant position is probably, as we have already noted, the widening use of tablets, e-readers, smartphones and the like. These typically do not use Microsoft operating systems. A new market has, therefore, provided a level playing field for firms to compete with Microsoft. ■

5. A Laissez-Faire Policy toward Natural Monopoly

As a fifth and final alternative for dealing with natural monopoly, we have the possibility of laissez-faire, or doing nothing—just letting the monopolist produce whatever quantity she chooses and sell it at whatever price the market will bear. The obvious objections to this policy are the two discussed in the previous section, namely, the efficiency and fairness problems. We may, however, decide in specific instances that these objections are not serious enough to warrant intervention.

This will in large part depend on whether the firm is producing an inferior, normal or luxury good. For instance, utilities like electricity, gas and water are clearly essential. To many this necessitates some form of policy intervention. We might, however, be less concerned that Legoland has a monopoly on Lego-themed amusement parks or that Real Madrid has a monopoly on, well, being Real Madrid. And we are probably even less concerned if Lamborghini or Bentley over-price their top-end models. Clearly, as we progress from inferior to luxury goods the fairness objection to monopoly becomes less pronounced because we move to a richer set of customers. The efficiency objective is also likely to diminish because the relevant market becomes that much smaller.

So what are we to conclude from this brief analysis of the five policy options for dealing with natural monopoly? The short answer is that each has problems. None completely eliminates the difficulties that arise when a single seller serves the market. Sometimes the least-costly solution will be competitive contracting, other times direct state ownership. Regulation will continue to play a role in specific industries, particularly the traditional public utilities. And despite their many shortcomings, antitrust laws serve the public well by discouraging price-fixing and other anticompetitive practices. But in some cases the best option may be simply not to intervene at all.

[16]See case number 37792 on the EU competition website (http://ec.europa.eu/competition).

PRICE DISCRIMINATION

Our discussion thus far has assumed that the monopolist sells all its output at a single price. In reality, however, monopolists often charge different prices to different buyers, a practice that is known as *price discrimination.* The cinema discount tickets discussed at the beginning of this chapter constitute one example. In this section we analyse how the profit-maximizing monopolist behaves when it is possible to charge different prices to different buyers. When price discrimination is possible, a monopolist can transfer some of the gains from consumers into its own profits. However, we will see that not all the higher profits under price discrimination come at the expense of consumers. Efficiency is enhanced as the monopolist expands output toward the level at which demand intersects marginal cost.

Third-Degree Price Discrimination

Suppose the monopolist has two completely distinct markets in which she can sell her output. Perhaps she is the only supplier in the domestic market for her product, and the only one in a foreign market as well. If she is a profit maximizer, what prices should she charge and what quantities should she sell in each market?

Suppose the demand and marginal revenue curves for the two markets are as given in the left and middle panels in Figure 13.17. First note that if the monopolist is maximizing profit, her marginal revenue should be the same in each market. (If it weren't, she could sell 1 unit less in the market with lower MR and 1 unit more in the market with higher MR, thereby increasing her profit.) Given that MR in the two markets must be the same, the profit-maximizing total quantity will be the one for which this common value is the same as marginal cost. Graphically, the solution is to add the marginal revenue curves horizontally across the two markets, and produce the level of output for which the resulting curve intersects the marginal cost curve. In the right panel in Figure 13.17, the optimal total output is indicated by $Q^* = 10$ units. $Q_1^* = 4$ of it is sold in market 1 at a price of P_1^*, and the remaining $Q_2^* = 6$ in market 2 at a price of P_2^*.

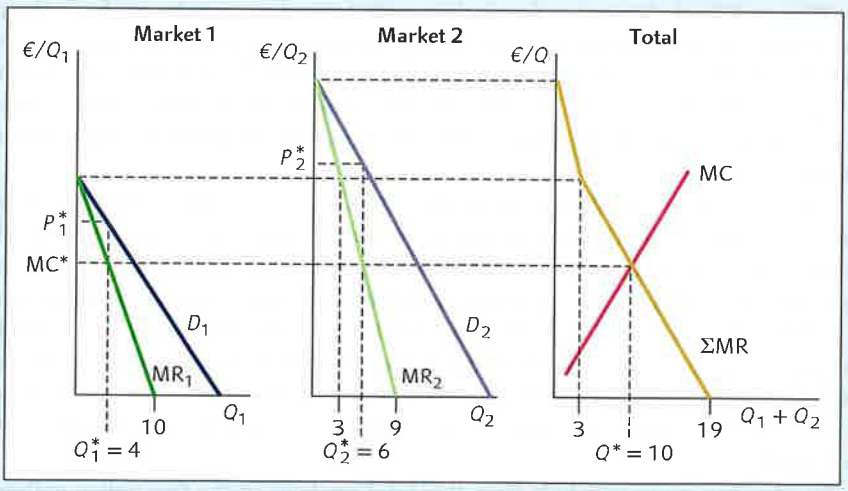


FIGURE 13.17
The Profit-Maximizing Monopolist Who Sells in Two Markets
For a monopolist who sells in two markets, the profit-maximizing output level is where the ΣMR curve intersects the MC curve, here, $Q^* = 10$. Marginal revenue in each market will be the same when $Q_1^* = 4$ and $Q_2^* = 6$ are sold in markets 1 and 2, respectively.

EXAMPLE 13.4 **A monopolist has marginal costs MC = Q and home market demand $P = 30 - Q$. The monopolist can also sell to a foreign market at a constant price $P_F = 12$. Find and graph the quantity produced, quantity sold in the home market, quantity sold in the foreign market and price charged in the home market. Explain why the monopolist's profits would fall if it were to produce the same quantity but sell more in the home market.**

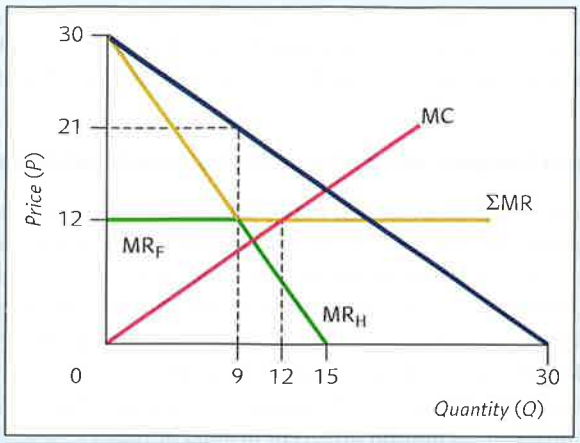


FIGURE 13.18
A Monopolist with a Perfectly Elastic Foreign Market
The curve ΣMR follows MR_H as long as $MR_H \geq MR_F$, and then follows MR_F. The profit-maximizing output level is where the ΣMR curve intersects the MC curve, here $Q^* = 12$.

The linear demand curve $P = 30 - Q$ has associated marginal revenue $MR = 30 - 2Q$. The profit-maximizing level of output for a monopolist selling to segmented markets occurs where $\Sigma MR = MC$. The horizontal sum of the marginal revenues across markets is the home marginal revenue function MR_H up to home output where $MR_F = MR_H$, and then the foreign marginal revenue function $MR_F = 12$ for any further units (see Figure 13.18). Total marginal revenue equals marginal cost at $MR_F = MC$, which solves for $Q = 12$. Marginal cost for this level of output equals home marginal revenue at $30 - 2Q_H = 12$, so $Q_H = 9$, with the remaining units sold abroad:

$$Q_F = Q - Q_H = 12 - 9 = 3$$

In the home market, the monopolist charges

$$P_H = 30 - Q_H = 30 - 9 = 21$$

Any further units sold at home would yield marginal revenue less than 12. Since sales to the foreign market yield a constant marginal revenue of 12, shifting sales to the home market would decrease profits due to the lost marginal revenue for each unit shifted. ◆

EXERCISE 13.5 **Suppose a monopolist sells in two separate markets, with demand curves given by $P_1 = 10 - Q_1$ and $P_2 = 20 - Q_2$, respectively. If her total cost curve is given by $TC = 5 + 2Q$, what quantities should she sell and what prices should she charge in the two markets?**

Note in Exercise 13.5 that the monopolist who sells in two markets charges a higher price in the market where demand is less elastic with respect to price. This result follows directly from Equation 13.6, which says that a profit-maximizing monopolist will set $MR = P[1 - (1/|\varepsilon|)]$. In the case of two markets, we set $MR_1 = MR_2$ to give

$$P_1\left(1 - \frac{1}{|\varepsilon_1|}\right) = P_2\left(1 - \frac{1}{|\varepsilon_2|}\right) \tag{13.8}$$

where $|\varepsilon_1|$ is the elasticity of demand in market 1 and $|\varepsilon_2|$ in market 2. If $|\varepsilon_1| > |\varepsilon_2|$ the monopolist should set $P_1 < P_2$.

Charging different prices to buyers in completely separate markets is often referred to as **third-degree price discrimination**. There is no special significance to the term 'third-degree' beyond the fact that this type of price discrimination happened to have been the third one that appeared in an early taxonomy.

third-degree price discrimination different prices are charged in different markets or to different categories of consumer.

By charging different prices in different markets the monopolist increases its profit. It will, therefore, look to distinguish markets wherever

possible. Typically, this involves distinguishing between different categories of buyer. For example, the university sports centre might distinguish between students, faculty and local residents. Examples of such discrimination abound. Any discount for students, old-age pensioners, children or local residents is an example of third-degree price discrimination.

ECONOMIC NATURALIST 13.6

Why do some doctors and lawyers offer discounts to people with low incomes?

In medicine, law, dentistry and other professions, many practitioners set their fees on a 'sliding scale'—in effect, selling their services to low-income consumers at significant discounts. This practice is often said to stem from professionals' concerns about the economic hardships confronting the poor. Such concerns are no doubt often heartfelt. But note also that the services offered by these professionals are normal goods, which means that the demand curves of low-income customers lie well below those of their wealthier counterparts. Sliding-scale fees may thus also be viewed as attempts by professionals to increase their profits by tailoring their prices to elasticity differences among different groups of buyers. A similar pattern is observed in the market for movie tickets, in which it is common for cinema owners to set lower prices for students, senior citizens and other groups believed to have higher price elasticities of demand. ■

Notice that third-degree price discrimination is feasible if two conditions hold. First, it needs to be possible to distinguish categories of consumer. For instance, to get a student discount it is necessary to show a student ID card. Furthermore, it must be impossible, or at least impractical, for buyers to trade among themselves. If, say, a textbook costs €20 for students and €30 for non-students then entrepreneurial students would buy €20 books and sell them for, say, €25 to non-students; others, hoping to get in on the action, would cut price even further, and eventually the price differential would vanish. Buying at a low price from one source and reselling at a higher price is often called **arbitrage**. Where arbitrage is practical, large price differentials for a single product cannot persist. Arbitrage ensures, for example, that the price of gold in London can never differ significantly from the price of gold in New York.

arbitrage the purchase of something for costless risk-free resale at a higher price.

ECONOMIC NATURALIST 13.7

Why do cinema owners offer student discounts on admission tickets but not on popcorn?

Arbitrage is practical in some cases but not in others. Student discounts on tickets enable cinemas to segment their markets because it is not possible for one person to see a movie at a low price and then sell the experience to someone else at a higher price. By the same token, it is practical for lawyers and doctors to charge different people different prices on the basis of differences in price elasticity of demand. But such market segmentation is more difficult for products like popcorn. If cinemas attempted to sell popcorn for €1 to students and for €3 to adults, some enterprising student would seize the arbitrage opportunity, selling popcorn to disgruntled adults for only €2. And under the pressure of competition from other arbitrageurs, the price differential would fall until the price differential was barely sufficient to make it worth the students' while to engage in the transaction. ■

The Perfectly Discriminating Monopolist

first-degree price discrimination consumers are charged individual prices that capture all consumer surplus.

Third-degree price discrimination allows the monopolist to increase profit. But, it is not the best it can do. **First-degree price discrimination** is the term used to describe the largest possible extent of discrimination. To understand one way in which this may work recall the example of two-part tariffs given in Chapter 6. We looked at a tennis club that charges an annual membership fee in addition to hourly court fees. To refresh your memory, consider Figure 13.19.

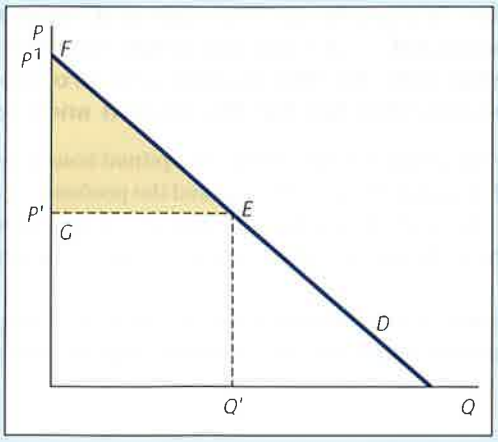


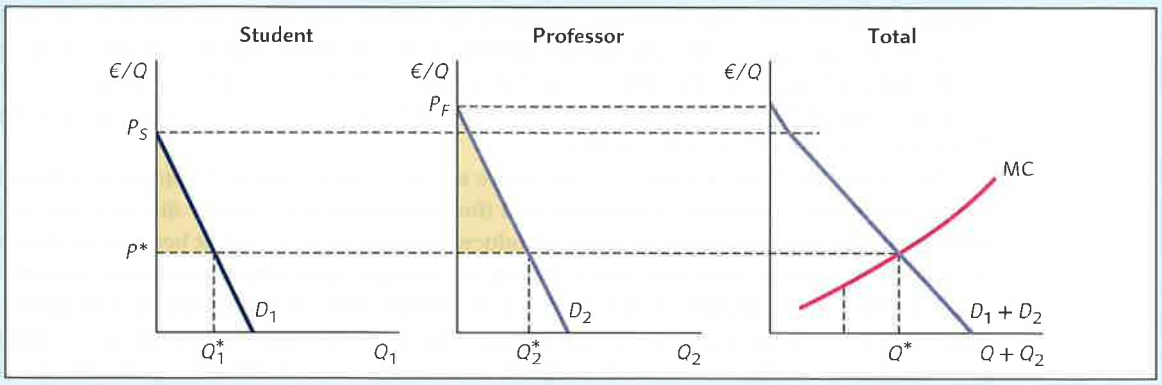
FIGURE 13.19
Two-Part Tariff
If the monopolist charges an hourly rate of P' and annual membership fee equal to the area EFG he can capture all consumer surplus.

This shows the demand curve of a particular consumer. With hourly court fees of P' the consumer would demand Q' hours of tennis and receive a consumer surplus equal to area EFG. This means the consumer is willing to pay an annual membership fee of $M = \frac{1}{2}\, Q'(P^1 - P')$. By charging an annual membership fee of M and hourly fee of P' the tennis club captures all consumer surplus.

So far, nothing new. What we need to do next is add into the mix other consumers and the costs of the tennis club. Suppose that there are two consumers—one a student and one a member of faculty. Because they carry a university ID card the tennis club can easily distinguish them. Figure 13.20 shows the individual and total demand curves, and the marginal cost curve. To find the optimal hourly fee we look to the total demand curve and the point where the marginal cost curve crosses the demand curve. It does so at price P^*. The tennis club should charge an hourly court fee of P^*. We then work out the maximum membership fee that each consumer is willing to pay. The student is willing to pay $M_s = \frac{1}{2} Q_1^* (P_S - P^*)$ and the professor is willing to pay $M_f = \frac{1}{2} Q_2^* (P_F - P^*)$.

By charging an hourly fee of P^* and annual membership fees of M_s and M_f the tennis club extracts the consumer surplus of both the student and professor. This is an example of first-degree price discrimination.

FIGURE 13.20
First-Degree Price Discrimination
If the monopolist charges an hourly rate of P^* he then charges the student and professor an annual membership fee equal to their respective consumer surplus. The hourly rate is the same for both consumers, but the membership fee is not.



EXAMPLE 13.5 **A tennis club has two customers. One is a student with demand curve $P = 10 - Q$. The other is a professor with demand curve $P = 20 - Q$. The tennis club has marginal costs MC = 2. Find and graph the optimal hourly court fee and annual membership fees for the student and professor. Explain why there is a difference in membership fee for the student and professor.**

We find the optimal hourly fee by setting $P = \text{MC}$. Thus, the optimal hourly fee is $P^* = 2$. With an hourly fee of €2 the student will demand 8 hours of tennis and the professor will demand 18 hours of tennis (see Figure 13.21). The student is willing to pay an annual membership fee of up to $M_s = \frac{1}{2}(8)(10 - 2) = €32$. The professor is willing to pay an annual membership fee of up to $M_f = \frac{1}{2}(18)(20 - 2) = €162$.

The professor pays a higher annual membership fee than the student because of his higher demand. The higher demand may reflect the professor's greater income or greater love of tennis. ◆

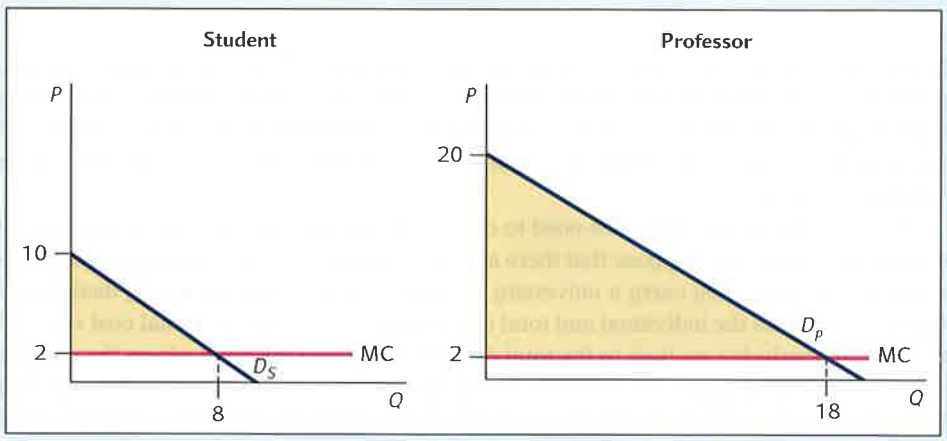


FIGURE 13.21
First-Degree Price Discrimination with Constant Marginal Cost
The monopolist can charge an hourly rate of $P^* = MC$. He then charges the student and professor an annual membership fee equal to their respective consumer surplus of €32 and €162.

First-degree price discrimination is characterized by the firm charging different consumers different prices *and* charging different prices for different units of output. To see how this works in the example, observe that the first hour of tennis costs the student $M_s + P^* = €34$ because he needs to pay both the annual membership and the hourly fee. The first hour of tennis costs the professor $M_f + P^* = €164$. The second hour of tennis costs both the student and professor only $P^* = €2$; the membership fee is a sunk cost by this point. By pricing in this way the tennis club captures all consumer surplus.

Two-part tariffs are not the only way to capture all consumer surplus. The tennis club could, for example, do away with the membership fee and charge each consumer the maximum they are willing to pay for each hour of tennis—charging the student, say, €9 for the first hour, €8 for the second hour, and so on. This will capture almost all the consumer surplus. Or, the tennis club could charge a bigger membership fee and scrap the hourly fee but limit use of the court. For example, it could charge the student a €48 membership fee and limit him to 8 hours of court time. This again captures all consumer surplus.

No matter what the monopolist does, there are two salient points of comparison between the perfectly discriminating monopolist and the monopolist who cannot discriminate at all. The first is that the perfect discriminator produces a higher level of output because he need not be concerned with the effect of a price cut on the revenue from output produced thus far. He can cut price to the people who would not otherwise buy, and maintain higher prices to those who are willing to pay them. And he can cut prices on some units of the good without lowering prices on others. Price and marginal revenue are one and the same, just as in the case of perfect competition.

A second important difference is that there generally is positive consumer surplus under the non-perfectly discriminating monopolist, but none under the perfect discriminator. With third-degree price discrimination, for instance, consumer surplus exists. If the monopolist must charge a buyer the same price for all units of the good then he cannot capture consumer surplus. Similarly, if he must charge the same price to all consumers the least elastic demanders end up paying a price well below their respective reservation prices—hence the consumer surplus.

Perfect price discrimination is a never-attained theoretical limit. If a customer's demand curve were tattooed on his forehead, it might be possible for a seller to tailor each price to extract the maximum possible amount from every buyer. But in general, the details of individual demand are only imperfectly known to the seller. The tennis club cannot charge each different consumer an individually tailored membership fee. Firms do, however, often estimate individual elasticity on the basis of information known about groups to which the individual belongs. Tennis clubs do charge different membership fees to different groups. This is a step towards perfect price discrimination.

Perhaps the closest thing we see to an in-depth assessment of individual elasticities is in the behaviour of merchants in bazaars in the Middle East. The shrewd camel trader has had many years of experience in trying to assess how much a buyer with a given demographic and psychological profile is willing to pay. His stock in trade is to interpret the incongruous gesture, the furtive eye movement. But even here, the wily buyer may know how to conceal his eagerness to own the camel.

Second-Degree Price Discrimination

Given that we have covered first- and third-degree price discrimination there are no prizes for guessing that second-degree price discrimination is next. But can you work out what characterizes second-degree price discrimination?

Third-degree price discrimination is where firms charge different prices in different markets, or to different categories of consumer. First-degree price discrimination is where firms charge different prices to different categories of consumer *and* charge different prices for different units of the good. **Second-degree price discrimination** is where firms charge difference prices for different units of the good.

second-degree price discrimination different quantities of the good sell for different prices.

Two-part tariffs are an example of second-degree price discrimination. Two-for-one offers are another example. Many electric utilities employ what are called *declining tail-block rate structures* by which the first, say, 300 kilowatt-hours per month are billed at 10 cents each, the next 700 at 8 cents, and all quantities over 1,000 kilowatt-hours/mo at 5 cents each. Such rate structures are also a form of second-degree price discrimination. Figure 13.22 illustrates the effect of such a rate structure for a consumer with the demand

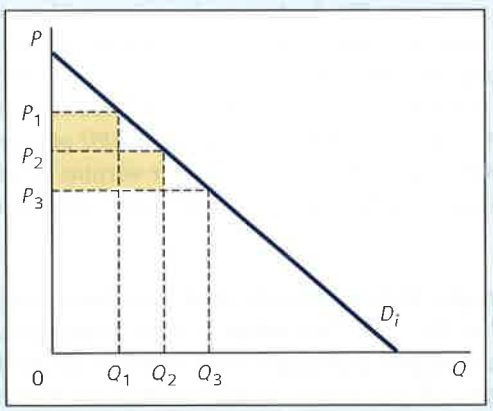


FIGURE 13.22
Second-Degree Price Discrimination
The seller offers the first block of consumption (0 to Q_1) at a high price (P_1), the second block (Q_1 to Q_2) at a lower price (P_2), the third block (Q_2 to Q_3) at a still lower price (P_3), and so on. This enables the monopolist to capture a substantial share of consumer surplus (the shaded area).

curve labelled D_i. In comparison with the alternative of charging a price of P_3 for every unit, the quantity discount scheme increases the consumer's total payment by an amount equal to the shaded area.

Second-degree price discrimination is distinguished from first-degree discrimination by the inability to charge different categories of consumer different prices. It is easy enough to tell apart student and faculty if they have a university ID card. It is less easy to tell apart 'rich' and 'poor' or 'tennis lovers' and 'occasional players'. And it is clearly not enough to just ask customers what category they fall into. If the tennis club announces a lower price, or lower membership fee, for 'poor' people then everyone clearly has an incentive to say they are poor.

The fact that the firm cannot directly distinguish categories does not mean, however, that they cannot do so indirectly. To illustrate, let us look again at the tennis club. Suppose that the club can no longer directly distinguish student from professor.

Figure 13.23 reproduces the demand curve of the student and professor in one diagram. The perfectly price-discriminating monopolist would charge an hourly fee P^* and a low membership fee M_s, equal to area *EFG*, for the student, and a high membership fee M_f, equal to area *EHJ*, for the professor. This pricing plan will not work if there is no way to distinguish the student from the professor. The professor would clearly prefer to pay the low fee and save some money. How can the tennis club avoid this?

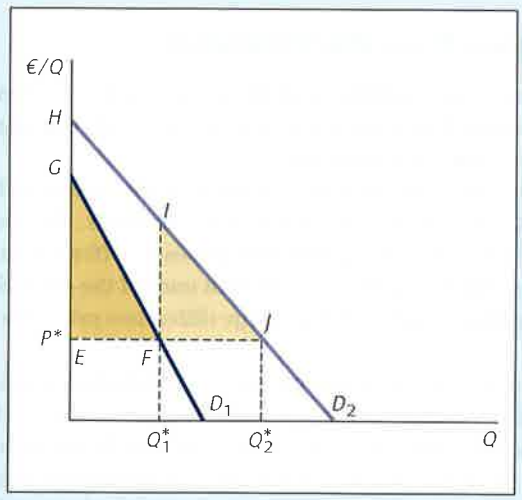


FIGURE 13.23
Second-Degree Price Discrimination with Two-Part Tariffs
The monopolist offers a budget membership package and allows users up to Q_1^* of court time. Each hour is charged at price P^*. The membership fee is equal to area *EFG*. If the professor chooses this package then he gets consumer surplus equal to area *FGHI*. The monopolist can offer an executive package that allows unlimited use and has membership fee equal to area *EFG* plus *FJI*.

What the tennis club can do is bundle together quantity and price. In particular, suppose the club offers two membership packages. The 'budget package' costs M_s and allows the member to play up to Q_1^* hours. The 'executive package' costs M' and allows unlimited use of the court. How high can M' be for the professor to choose the executive package? If the professor chooses the budget package he is limited to Q_1^* hours and so his consumer surplus is equal to area *HGFI*. That means the professor will be willing to pay area *EFG* plus area *FJI* to get the executive package. He would then have the same consumer surplus whether he buys the budget or executive package. In this way the tennis club can extract some, if not all, consumer surplus from the professor.

EXAMPLE 13.6 **As in Example 13.5, a tennis club has two customers. One is poor with demand curve $P = 10 - Q$. The other is rich with demand curve $P = 20 - Q$. The club offers a budget package that costs €32 and allows up to 8 hours of court time at cost €2 per hour. It also offers an executive package that costs €M and allows unlimited court time at cost €2 per hour. What is the**

maximum the rich customer will pay for the executive package? How does this amount compare to the fee that can be charged if the club can distinguish between the rich and poor customer?

The budget package is clearly aimed towards the poor customer. If the rich consumer buys the budget package then his consumer surplus will equal area *FGHI* (Figure 13.24). This works out at (10)(8) = €80. If the executive package costs an amount equal to area *EFG* plus *FIJ* he would be indifferent between the executive and budget packages. This gives a membership fee of $M = €32 + \frac{1}{2}(18 - 8)(12 - 2) = €82$. At this price the rich consumer still gets consumer surplus equal to area *FGHI* if he buys the executive package.

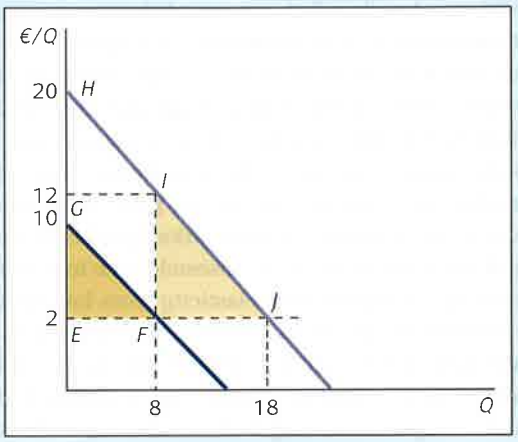


FIGURE 13.24
Second-Degree Price Discrimination Example
The monopolist charges an hourly rate of €2. The budget package costs €32, area *EFG*, and limits use to 8 hours. If the rich consumer were to buy this package his consumer surplus would be €80, area *GHIF*. He is willing to pay €82, areas *EFG* plus *FIJ*, to get an executive package offering unlimited use of the court.

We saw in Example 13.5 that the rich consumer would be willing to pay a membership fee of €162 if he is not allowed to buy the budget package. If he can buy the budget package he is no longer willing to pay €162 for the executive package—he would switch to the budget package. Consequently the fee is set at only €82 and the rich consumer gets a consumer surplus of €80. ◆

Bundling quantity and price together like this means different units of the good cost different prices. In the previous example, it costs (8)(2) + 32 = €48 to use the court for 8 hours, or €6 per hour. It costs (18)(2) + 82 = €118 to use the court for 18 hours, or €6.56 per hour. Because different units of the good cost different prices it is second-degree price discrimination, even though the objective is partly to distinguish between different categories of customer as in third-degree price discrimination.

Bundling quantity and price together is very common. Mobile phone packages are one obvious example. In this case the number of calls, texts or downloads are bundled into differently priced packages from pay as you go to unlimited use. Builders merchants also typically bundle together quantity and price to distinguish professional, trade customers from domestic, DIY customers.

mihailomilovanovic

Someone buying a new smartphone must choose between an array of differently priced models, options and packages.

The Hurdle Model of Price Discrimination

Bundling together price and quantity is not the only possibility open to the monopolist. It can also bundle together price and quality. In this case price discrimination is being used as a technique whereby

the firm induces the most elastic buyers to identify themselves. This is the *hurdle model of price discrimination.* The basic idea is that the seller sets up a hurdle of some sort and makes a discount price available to those buyers who elect to jump over it. The logic is that those buyers who are most sensitive to price will be more likely than others to jump the hurdle.

One example of a hurdle is a rebate form included in the product package. Here, jumping over the hurdle means filling in the form, finding a stamp and an envelope, and then getting to the post box to mail it in. The firm's hope is that people who do not care much about price will be less likely than others to bother going through this process. If so, then people whose demands are less elastic end up paying the 'regular' price, while those with more elastic demands pay the lower discount price.

It is a rare product whose seller does not use the hurdle model of differential pricing. Booksellers offer only high-priced hardback editions in the first year of publication. Buyers who do not care strongly about price buy these editions when they first come out. Others wait a year or two and then buy the much less expensive softcover edition. Here, the hurdle is having to endure the wait. Airlines offer 'super-saver' discounts of up to half off the regular standard fare. Here also there are two common hurdles: having to make reservations a week or more in advance and having to stay over a Saturday night. Many retailers include discount coupons in their newspaper ads. Here, the hurdles are having to read the ads, clip the coupons, and get to the store before they expire. Some sellers post signs behind the counter saying 'Ask about our special low price'. Here, the hurdle is merely having to do the asking. But even this trivial hurdle can be remarkably effective, because many well-heeled buyers would find asking about a special price too unseemly even to contemplate.

None of these schemes perfectly segregates high-elasticity from low-elasticity buyers. For instance, there are some people who wait for the January sales to buy their towels even though they would buy just as many if the sales were not offered. But on the whole, the hurdles seem to function much as intended. A perfect hurdle would be one that imposes only a negligible cost on the buyers who jump it, yet perfectly separates buyers according to their elasticity of demand. Analytically, the effect of such a hurdle is portrayed in Figure 13.25, where P_H represents the 'regular' price and P_L represents the discount price. With a perfect hurdle, none of the people who pay the discount price has a reservation price greater than or equal to the regular price, which means that all of them would have been excluded from the market had only the regular price been available.

The hurdle model need not be limited to the two-price version depicted in Figure 13.25. On the contrary, many sellers have developed it into a highly complex art form involving literally dozens of price–hurdle combinations. On its Amsterdam–New York route alone, for example, KLM offers dozens of different fares, each with its own set of restrictions. But no matter how simple or complex the scheme may be, its goal is the same—to give discounts to customers who would not otherwise buy the product.

The hurdle model is like first-degree price discrimination in that it tries to tailor prices to the elasticities of individual buyers. The principal difference is that even in its most sophisticated form, the hurdle model cannot hope to capture all the consumer surplus.

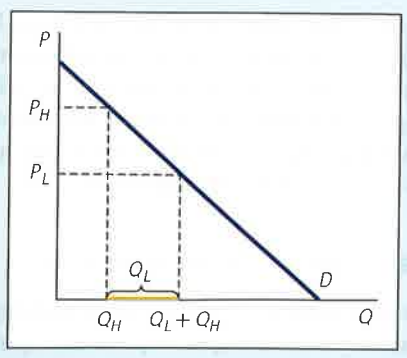


FIGURE 13.25
A Perfect Hurdle
When a hurdle is perfect, the only buyers who become eligible for the discount price (P_L) by jumping it are those who would not have been willing to pay the regular price (P_H). A perfect hurdle also imposes no significant costs on those who jump it.

Efficiency of a Price-Discriminating Monopolist

Consider a natural monopolist who uses the hurdle model of differential pricing. To keep the discussion simple, let's suppose she charges a regular price and also a discount price, the latter available to customers who clear some hurdle, such as mailing in a rebate form. How does the presence of this differential pricing device affect the efficiency and fairness and objections to monopoly?

Consider first the efficiency objection. Recall that the problem is that the single-price monopolist charges a price above marginal cost, which excludes many potential buyers from the market, ones who value the product more highly than the value of the resources required to produce it. For the example depicted in panel (a) of Figure 13.26 this deadweight loss is equal to area W.

Suppose instead that the monopolist is able to charge one price to the buyers along the upper part of the demand curve and a lower price to all other buyers (panel (*b*) in Figure 13.26). Suppose her profit-maximizing strategy will be to sell Q_H at the price P_H and Q_L at the price P_L.[17] The efficiency loss associated with the two-price monopolist is only area Z, which is much smaller than the corresponding loss for the single-price monopolist.

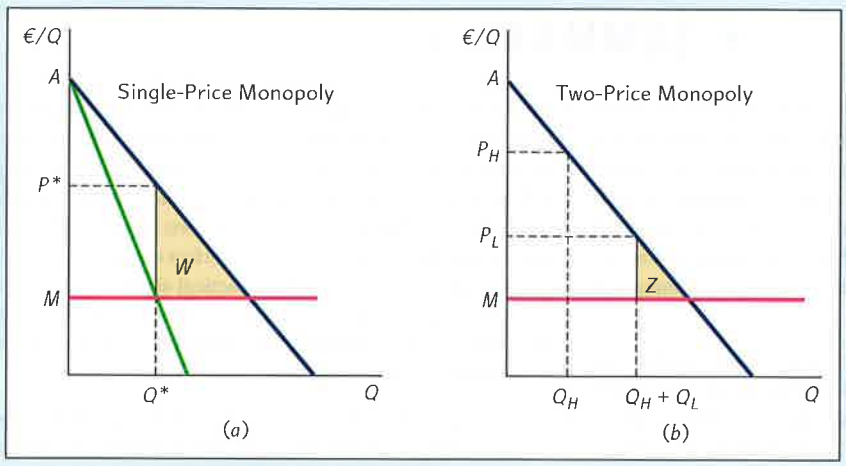


FIGURE 13.26
The Efficiency Losses from Single-Price and Two-Price Monopoly
By being able to offer a discount price to the most elastic portion of the demand curve, the two-price monopolist (panel *b*) expands the market, thereby causing a much smaller efficiency loss (area *Z*, panel *b*) than in the case of the single-price monopolist (area *W*, panel *a*).

[17]For the single-price monopolist the profit function is given by

$$\Pi_1 = (A - BQ)Q - F - MQ$$

The first-order condition for a maximum is given by

$$\frac{d\Pi_1}{dQ} = A - 2BQ - M = 0$$

which yields a profit-maximizing quantity of $Q' = (A - M)/2B$, and a corresponding price of $P' = (A + M)/2$. The profit function for the two-price monopolist, by contrast, is given by

$$\Pi_2 = (A - BQ_H)Q_H + (A - BQ_H - BQ_L)Q_L - F - M(Q_H + Q_L)$$

The first-order conditions for a maximum are given by

$$\frac{\partial \Pi_2}{\partial Q_H} = A - 2BQ_H - BQ_L - M = 0$$

and

$$\frac{\partial \Pi_2}{\partial Q_L} = A - BQ_H - 2BQ_L - M = 0$$

which can be solved for

$$Q_H = \frac{A - M}{3B} = Q_L \text{ and } P_L = \frac{A + 2M}{3} \text{ and } P_H = \frac{2A + M}{3}$$

In general, the more finely the monopolist can partition her market, the smaller the efficiency loss will be. Indeed, if it were possible for the monopolist to charge different prices to every buyer output would expand to the efficient level. As noted earlier, it is common in most firms to see not one but a whole menu of different discount prices, each with a different set of restrictions (the deeper the discount, the more stringent the restriction). Given the wide latitude many firms have to expand their markets through price discrimination, the efficiency problem of monopoly will often be of only secondary importance.

Consider next the fairness objection to monopoly. Recall that the problem is that the monopolist extracts consumer surplus. On this score price discrimination has mixed results. On the one hand price discrimination may make a product available to consumers who would not have purchased it otherwise. *Some* consumers may, therefore, benefit from differential pricing. The firm's objective in price discrimination is, however, to increase profit by extracting more consumer surplus. We know, therefore, that differential pricing must result in the monopolist extracting more *overall* consumer surplus; otherwise there would be no point in differential pricing. So, price discrimination makes the fairness objection to monopoly more acute.

SUMMARY

- Monopoly is the name given to the market structure in which a single firm serves the entire market. Five factors, acting alone or in combination, give rise to monopoly: (1) control over key inputs, (2) economies of scale, (3) patents, (4) network economies, and (5) government licences. In the long run, by far the most important of these is economies of scale, in part because it also helps explain network economies and government licences.
- Because the monopolist is the only seller in the market, his demand curve is the downward-sloping market demand curve. Unlike the perfect competitor, who can sell as much as he chooses at the market price, the monopolist must cut price in order to expand his output. The monopolist's rule for maximizing profits is the same as the one used by perfectly competitive firms. It is to expand output if the gain in revenue (marginal revenue) exceeds the increase in costs (marginal cost), and to contract if the loss in revenue is smaller than the reduction in costs. The pivotal difference is that marginal revenue is less than price for the monopolist, but equal to price for the perfect competitor.
- When the monopolist can sell in several separate markets, he distributes output among them so that marginal revenue is the same in each. Here again, the familiar logic of cost–benefit analysis provides a convenient framework for analysing the firm's decision about whether to alter its current behaviour.
- Unlike the perfectly competitive case, the monopoly equilibrium generally does not exhaust the potential gains from exchange. In general, the value to society of an additional unit of output will exceed the cost to the monopolist of the resources required to produce it. This finding has often been interpreted to mean that monopoly is less efficient than perfect competition. But this interpretation is of only limited practical significance, because the conditions that give rise to monopoly—in particular, economies of scale in production—are rarely compatible with those required for perfect competition.
- Our policy focus in the chapter was on the question of how the government should treat natural monopolies—markets characterized by downward-sloping long-run average cost curves. We considered five policy alternatives: (1) state ownership, (2) private ownership with government price regulation, (3) competitive bidding by private firms for the right to be the sole provider of service, (4) vigorous enforcement of antitrust laws designed to prevent monopoly, and finally (5) a complete laissez-faire, or hands-off, policy. Problems arise with each of these alternatives, and the best policy will in general be different in different circumstances. The laissez-faire stance is most attractive in markets where the monopolist is able to employ the hurdle model of differential pricing. Allowing buyers to decide for themselves whether to become eligible for a discount price softens both the efficiency and fairness objections to natural monopoly.

QUESTIONS FOR REVIEW

1. What five factors give rise to monopoly? In the long run, why are economies of scale the most important factor?
2. If the United States has thousands of cement producers but a small town has only one, is this cement producer a monopolist? Explain.