CHAPTER

8

Firms in the Global Economy: Export Decisions, Outsourcing, and Multinational Enterprises

I n this chapter, we continue to explore how economies of scale generate incentives for international specialization and trade. We now focus on economies of scale that are internal to the firm. As mentioned in the previous chapter, this form of increasing returns leads to a market structure that features imperfect competition. **Internal economies of scale** imply that a firm's average cost of production decreases the more output it produces. Perfect competition that drives the price of a good down to marginal cost would imply losses for those firms because they would not be able to recover the higher costs incurred from producing the initial units of output.[1] As a result, perfect competition would force those firms out of the market, and this process would continue until an equilibrium featuring imperfect competition is attained.

Modeling imperfect competition means that we will explicitly consider the behavior of individual firms. This will allow us to introduce two additional characteristics of firms that are prevalent in the real world: (1) In most sectors, firms produce goods that are differentiated from one another. In the case of certain goods (such as bottled water, staples, etc.), those differences across products may be small, while in others (such as cars, cell phones, etc.), the differences are much more significant. (2) Performance measures (such as size and profits) vary widely across firms. We will incorporate this first characteristic (product differentiation) into our analysis throughout this chapter. To ease exposition and build intuition, we will initially consider the case when there are no performance differences between firms. We will thus see how internal economies of scale and product differentiation combine to generate some new sources of gains of trade via economic integration.

We will then introduce differences across firms so that we can analyze how firms respond differently to international forces. We will see how economic

---

[1]Whenever average cost is decreasing, the cost of producing one extra unit of output (marginal cost) is lower than the average cost of production (since that average includes the cost of those initial units that were produced at higher unit costs).

integration generates both winners and losers among different types of firms. The better-performing firms thrive and expand, while the worse-performing firms contract. This generates one additional source of gain from trade: As production is concentrated toward better-performing firms, the overall efficiency of the industry improves. Lastly, we will study why those better-performing firms have a greater incentive to engage in the global economy, either by exporting, by outsourcing some of their intermediate production processes abroad, or by becoming multinationals and operating in multiple countries.

LEARNING GOALS

After reading this chapter, you will be able to:

- Understand how internal economies of scale and product differentiation lead to international trade and intra-industry trade.
- Recognize the new types of welfare gains from intra-industry trade.
- Describe how economic integration can lead to both winners and losers among firms in the same industry.
- Explain why economists believe that "dumping" should not be singled out as an unfair trade practice, and why the enforcement of antidumping laws leads to protectionism.
- Explain why firms that engage in the global economy (exporters, outsourcers, multinationals) are substantially larger and perform better than firms that do not interact with foreign markets.
- Understand theories that explain the existence of multinationals and the motivation for foreign direct investment across economies.

The Theory of Imperfect Competition

In a perfectly competitive market—a market in which there are many buyers and sellers, none of whom represents a large part of the market—firms are *price takers.* That is, they are sellers of products who believe they can sell as much as they like at the current price but cannot influence the price they receive for their product. For example, a wheat farmer can sell as much wheat as she likes without worrying that if she tries to sell more wheat, she will depress the market price. The reason she need not worry about the effect of her sales on prices is that any individual wheat grower represents only a tiny fraction of the world market.

When only a few firms produce a good, however, the situation is different. To take perhaps the most dramatic example, the aircraft manufacturing giant Boeing shares the market for large jet aircraft with only one major rival, the European firm Airbus. As a result, Boeing knows that if it produces more aircraft, it will have a significant effect on the total supply of planes in the world and will therefore significantly drive down the price of airplanes. Or to put it another way, Boeing knows that if it wants to sell more airplanes, it can do so only by significantly reducing its price. In **imperfect competition,** then, firms are aware that they can influence the prices of their products and that they can sell more only by reducing their price. This situation occurs in one of two ways: when there are only a few major producers of a particular good, or when each firm produces a good that is differentiated (in the eyes of the consumer) from that of rival firms. As we mentioned in the introduction, this type of competition is an inevitable outcome when there are economies

of scale at the level of the firm: The number of surviving firms is forced down to a small number and/or firms must develop products that are clearly differentiated from those produced by their rivals. Under these circumstances, each firm views itself as a *price setter*, choosing the price of its product, rather than a price taker.

When firms are not price takers, it is necessary to develop additional tools to describe how prices and outputs are determined. The simplest imperfectly competitive market structure to examine is that of a **pure monopoly,** a market in which a firm faces no competition; the tools we develop for this structure can then be used to examine more complex market structures.

Monopoly: A Brief Review

Figure 8-1 shows the position of a single monopolistic firm. The firm faces a downward-sloping demand curve, shown in the figure as *D*. The downward slope of *D* indicates that the firm can sell more units of output only if the price of the output falls. As you may recall from basic microeconomics, a **marginal revenue** curve corresponds to the demand curve. Marginal revenue is the extra or marginal revenue the firm gains from selling an additional unit. Marginal revenue for a monopolist is always less than the price because to sell an additional unit, the firm must lower the price of *all* units (not just the marginal one). Thus for a monopolist, the marginal revenue curve, *MR*, always lies below the demand curve.

Marginal Revenue and Price   For our analysis of the monopolistic competition model later in this section, it is important for us to determine the relationship between the price the monopolist receives per unit and marginal revenue. Marginal revenue is always less than the price—but how much less? The relationship between marginal revenue and price depends on two things. First, it depends on how much output the firm is already selling: A firm that is not selling very many units will not lose much by cutting the price it receives on those units. Second, the gap between price and marginal revenue depends on the slope of the demand curve, which tells us how much the monopolist has to cut his price to sell one more unit of output. If the curve is very flat, then the monopolist can sell an additional unit with only a small price cut. As a result, he will not have to lower the price by very

Figure 8-1

Monopolistic Pricing and Production Decisions

A monopolistic firm chooses an output at which marginal revenue, the increase in revenue from selling an additional unit, equals marginal cost, the cost of producing an additional unit. This profit-maximizing output is shown as Q_M; the price at which this output is demanded is P_M. The marginal revenue curve *MR* lies below the demand curve *D* because, for a monopoly, marginal revenue is always less than the price. The monopoly's profits are equal to the area of the shaded rectangle, the difference between price and average cost times the amount of output sold.
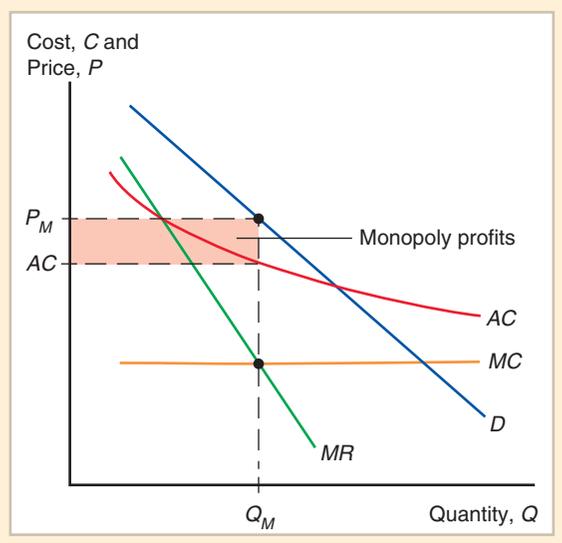
much on the units he would otherwise have sold, so marginal revenue will be close to the price per unit. On the other hand, if the demand curve is very steep, selling an additional unit will require a large price cut, implying that marginal revenue will be much less than the price.

We can be more specific about the relationship between price and marginal revenue if we assume that the demand curve the firm faces is a straight line. When this is the case, the dependence of the monopolist's total sales on the price it charges can be represented by an equation of the form

$$Q = A - B \times P, \tag{8-1}$$

where Q is the number of units the firm sells, P the price it charges per unit, and A and B are constants. We show in the appendix to this chapter that in this case, marginal revenue is

$$\text{Marginal revenue} = MR = P - Q/B, \tag{8-2}$$

implying that

$$P - MR = Q/B.$$

Equation (8-2) reveals that the gap between price and marginal revenue depends on the initial sales, Q, of the firm and the slope parameter, B, of its demand curve. If sales quantity, Q, is higher, marginal revenue is lower, because the decrease in price required to sell a greater quantity costs the firm more. In other words, the greater is B, the more sales fall for any given increase in price and the closer the marginal revenue is to the price of the good. Equation (8-2) is crucial for our analysis of the monopolistic competition model of trade in the upcoming section.

Average and Marginal Costs   Returning to Figure 8-1, AC represents the firm's **average cost** of production, that is, its total cost divided by its output. The downward slope reflects our assumption that there are economies of scale, so the larger the firm's output, the lower its costs per unit. MC represents the firm's **marginal cost** (the amount it costs the firm to produce one extra unit). In the figure, we assumed that the firm's marginal cost is constant (the marginal cost curve is flat). The economies of scale must then come from a fixed production cost. This fixed cost pushes the average cost above the constant marginal cost of production, though the difference between the two becomes smaller and smaller as the fixed cost is spread over an increasing number of output units.

If we denote c as the firm's marginal cost and F as the fixed cost, then we can write the firm's total cost (C) as

$$C = F + c \times Q, \tag{8-3}$$

where Q is once again the firm's output. Given this linear cost function, the firm's average cost is

$$AC = C/Q = (F/Q) + c. \tag{8-4}$$

As we have discussed, this average cost is always greater than the marginal cost c, and declines with output produced Q.

If, for example, $F = 5$ and $c = 1$, the average cost of producing 10 units is $(5/10) + 1 = 1.5$, and the average cost of producing 25 units is $(5/25) + 1 = 1.2$. These numbers may look familiar, because they were used to construct Table 7-1 in the

Figure 8-2

Average versus Marginal Cost

This figure illustrates the average and marginal costs corresponding to the total cost function $C = 5 + x$. Marginal cost is always 1; average cost declines as output rises.
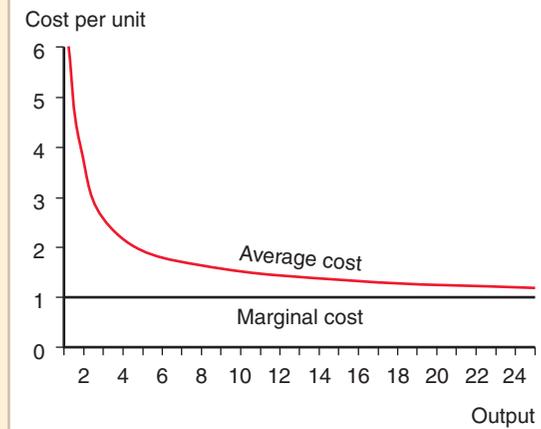
previous chapter. (However, in this case, we assume a unit wage cost for the labor input, and that the technology now applies to a firm instead of an industry.) The marginal and average cost curves for this specific numeric example are plotted in Figure 8-2. Average cost approaches infinity at zero output and approaches marginal cost at very large output.

The profit-maximizing output of a monopolist is that at which marginal revenue (the revenue gained from selling an extra unit) equals marginal cost (the cost of producing an extra unit), that is, at the intersection of the *MC* and *MR* curves. In Figure 8-1 we can see that the price at which the profit-maximizing output Q_M is demanded is P_M, which is greater than average cost. When $P > AC$, the monopolist is earning some monopoly profits, as indicated by the shaded box.[2]

Monopolistic Competition

Monopoly profits rarely go uncontested. A firm making high profits normally attracts competitors. Thus situations of pure monopoly are rare in practice. Instead, the usual market structure in industries characterized by internal economies of scale is one of **oligopoly,** in which several firms are each large enough to affect prices, but none has an uncontested monopoly.

The general analysis of oligopoly is a complex and controversial subject because in oligopolies, the pricing policies of firms are *interdependent.* Each firm in an oligopoly will, in setting its price, consider not only the responses of consumers but also the expected responses of competitors. These responses, however, depend in turn on the competitors' expectations about the firm's behavior—and we are therefore in a complex game in which firms are trying to second-guess each other's strategies. We will briefly discuss an example of an oligopoly model with two firms in Chapter 12. For now, we focus on a special case of oligopoly known as **monopolistic competition**. Over the last 30 years, research in

---

[2]The economic definition of *profits* is not the same as that used in conventional accounting, where any revenue over and above labor and material costs is called a profit. A firm that earns a rate of return on its capital less than what that capital could have earned in other industries is not making profits; from an economic point of view, the normal rate of return on capital represents part of the firm's costs, and only returns over and above that normal rate of return represent profits.

international trade has increasingly relied on models based on monopolistic competition. This model can capture the key elements of imperfect competition based on internal economies of scale and product differentiation at the firm level. At the same time, this model remains relatively easy to analyze, even in a setting where economy-wide prices are affected by international trade.

In monopolistic competition models, two key assumptions are made to get around the problem of interdependence. First, each firm is assumed to be able to *differentiate its product* from that of its rivals. That is, because a firm's customers want to buy that particular firm's product, they will not rush to buy other firms' products because of a slight price difference. Product differentiation thus ensures that each firm has a monopoly in its particular product within an industry and is therefore somewhat insulated from competition. Second, each firm is assumed to take the prices charged by its rivals as given—that is, it ignores the impact of its own price on the prices of other firms. As a result, the monopolistic competition model assumes that even though each firm is in reality facing competition from other firms, each firm behaves as if it were a monopolist—hence the model's name.

Are there any monopolistically competitive industries in the real world? The first assumption of product differentiation across firms fits very well with the empirical evidence in most industries. The extent of product differentiation varies widely across industries, but consumers do perceive differences across products sold by different firms in most sectors (even if the "actual" differences across products are very small, such as in the case of bottled water). The second assumption—that firms ignore the consequence on rival firms of their pricing decisions—is more of an approximation. In some sectors (such as large jet aircraft), a small number of firms account for a very large percentage of the overall market share. Firms in those sectors are much more likely to engage in strategic pricing decisions with their rivals. However, these strategic effects dissipate quickly as the market share of the largest firms drops. In any event, the main appeal of the monopolistic competition model is not its realism but its simplicity. As we will see in the next section of this chapter, the monopolistic competition model gives us a very clear view of how economies of scale can give rise to mutually beneficial trade.

Before we can examine trade, however, we need to develop a basic model of monopolistic competition. Let us therefore imagine an industry consisting of a small number of firms. These firms produce differentiated products, that is, goods that are not exactly the same but that could be substitutes for one another. Each firm is therefore a monopolist in the sense that it is the only firm producing its particular good, but the demand for its good depends on the number of other similar products available and on the prices of other firms' products in the industry.

Assumptions of the Model   We begin by describing the demand facing a typical monopolistically competitive firm. In general, we would expect a firm to sell more the larger the total demand for its industry's product and the higher the prices charged by its rivals. On the other hand, we would expect the firm to sell less the greater the number of firms in the industry and the higher its own price. A particular equation for the demand facing a firm that has these properties is[3]

$$Q = S \times [1/n - b \times (P - \bar{P})], \tag{8-5}$$

[3]Equation (8-5) can be derived from a model in which consumers have different preferences and firms produce varieties tailored to particular segments of the market. See Stephen Salop, "Monopolistic Competition with Outside Goods," *Bell Journal of Economics* 10 (1979), pp. 141–156, for a development of this approach.

where Q is the quantity of output demanded, S is the total output of the industry, n is the number of firms in the industry, b is a constant term representing the responsiveness of a firm's sales to its price, P is the price charged by the firm itself, and $\overline{P}$ is the average price charged by its competitors. Equation (8-5) may be given the following intuitive justification: If all firms charge the same price, each will have a market share $1/n$. A firm charging more than the average of other firms will have a smaller market share, whereas a firm charging less will have a larger share.[4]

It is helpful to assume that total industry output S is unaffected by the average price $\overline{P}$ charged by firms in the industry. That is, we assume that firms can gain customers only at each other's expense. This is an unrealistic assumption, but it simplifies the analysis and helps us focus on the competition among firms. In particular, it means that S is a measure of the size of the market and that if all firms charge the same price, each sells S/n units.

Next we turn to the costs of a typical firm. Here we simply assume that total and average costs of a typical firm are described by equations (8-3) and (8-4). Note that in this initial model, we assume that all firms are *symmetric* even though they produce differentiated products: They all face the same demand curve (8-5) and have the same cost function (8-3). We will relax this assumption in the next section.

Market Equilibrium   When the individual firms are symmetric, the state of the industry can be described without describing any of the features of individual firms: All we really need to know to describe the industry is how many firms there are and what price the typical firm charges. To analyze the industry—for example, to assess the effects of international trade—we need to determine the number of firms n and the average price they charge $\overline{P}$. Once we have a method for determining n and $\overline{P}$, we can ask how they are affected by international trade.

Our method for determining n and $\overline{P}$ involves three steps. (1) First, we derive a relationship between the number of firms and the *average cost* of a typical firm. We show that this relationship is upward sloping; that is, the more firms there are, the lower the output of each firm, and thus the higher each firm's cost per unit of output. (2) We next show the relationship between the number of firms and the price each firm charges, which must equal $\overline{P}$ in equilibrium. We show that this relationship is downward sloping: The more firms there are, the more intense is the competition among firms, and as a result the lower the prices they charge. (3) Finally, we introduce firm entry and exit decisions based on the profits that each firm earns. When price exceeds average cost, firms earn positive profits and additional firms will enter the industry; conversely, when the price is less than average cost, profits are negative and those losses induce some firms to exit. In the long run, this entry and exit process drives profits to zero, and the number of firms is determined by the intersection of the curve that relates average cost to n and the curve that relates price to n.

1. *The number of firms and average cost.* As a first step toward determining n and $\overline{P}$, we ask how the average cost of a typical firm depends on the number of firms in the industry. Since all firms are symmetric in this model, in equilibrium they all will charge the same price. But when all firms charge the same price, so that $P = \overline{P}$, equation (8-5) tells us that $Q = S/n$; that is, each firm's output Q is a $1/n$ share of the total industry sales S. But we saw in equation (8-4) that average cost depends inversely

on a firm's output. We therefore conclude that average cost depends on the size of the market and the number of firms in the industry:

$$AC = F/Q + c = (n \times F/S) + c. \qquad \text{(8-6)}$$

Equation (8-6) tells us that other things equal, *the more firms there are in the industry, the higher is average cost.* The reason is that the more firms there are, the less each firm produces. For example, imagine an industry with total sales of 1 million widgets annually. If there are five firms in the industry, each will sell 200,000 annually. If there are ten firms, each will sell only 100,000, and therefore each firm will have higher average cost. The upward-sloping relationship between n and average cost is shown as CC in Figure 8-3.

2. *The number of firms and the price.* Meanwhile, the price the typical firm charges also depends on the number of firms in the industry. In general, we would expect that the more firms there are, the more intense will be the competition among them, and
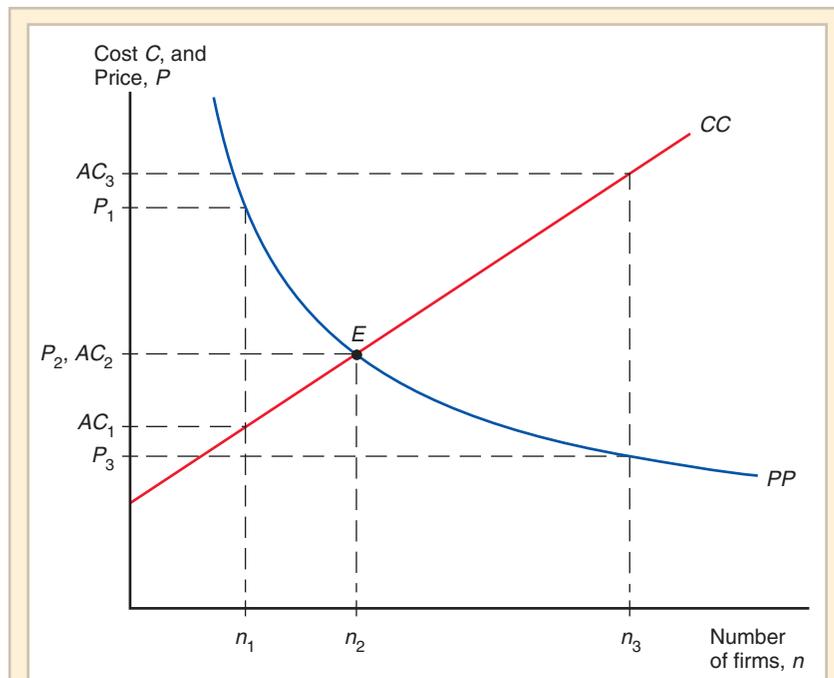
Figure 8-3

Equilibrium in a Monopolistically Competitive Market

The number of firms in a monopolistically competitive market, and the prices they charge, are determined by two relationships. On one side, the more firms there are, the more intensely they compete, and hence the lower is the industry price. This relationship is represented by *PP*. On the other side, the more firms there are, the less each firm sells and therefore the higher is the industry's average cost. This relationship is represented by *CC*. If price exceeds average cost (that is, if the *PP* curve is above the *CC* curve), the industry will be making profits and additional firms will enter the industry; if price is less than average cost, the industry will be incurring losses and firms will leave the industry. The equilibrium price and number of firms occurs when price equals average cost, at the intersection of *PP* and *CC*.

hence the lower the price. This turns out to be true in this model, but proving it takes a moment. The basic trick is to show that each firm faces a straight-line demand curve of the form we showed in equation (8-1), and then to use equation (8-2) to determine prices.

First recall that in the monopolistic competition model, firms are assumed to take each other's prices as given; that is, each firm ignores the possibility that if it changes its price, other firms will also change theirs. If each firm treats $\bar{P}$ as given, we can rewrite the demand curve (8-5) in the form

$$Q = [(S/n) + S \times b \times \bar{P}] - S \times b \times P, \tag{8-7}$$

where b is the parameter in equation (8-5) that measured the sensitivity of each firm's market share to the price it charges. Now this equation is in the same form as (8-1), with $(S/n) + S \times b \times \bar{P}$ in place of the constant term A and $S \times b$ in place of the slope coefficient B. If we plug these values back into the formula for marginal revenue, (8-2), we have a marginal revenue for a typical firm of

$$MR = P - Q/(S \times b). \tag{8-8}$$

Profit-maximizing firms will set marginal revenue equal to their marginal cost, c, so that

$$MR = P - Q/(S \times b) = c,$$

which can be rearranged to give the following equation for the price charged by a typical firm:

$$P = c + Q/(S \times b). \tag{8-9}$$

We have already noted, however, that if all firms charge the same price, each will sell an amount $Q = S/n$. Plugging this back into (8-9) gives us a relationship between the number of firms and the price each firm charges:

$$P = c + 1/(b \times n). \tag{8-10}$$

Equation (8-10) says algebraically that *the more firms there are in an industry, the lower the price each firm will charge.* This is because each firm's **markup over marginal cost,** $P - c = 1/(b \times n)$, decreases with the number of competing firms. Equation (8-10) is shown in Figure 8-3 as the downward-sloping curve *PP*.

3. *The equilibrium number of firms.* Let us now ask what Figure 8-3 means. We have summarized an industry by two curves. The downward-sloping curve *PP* shows that the more firms there are in the industry, the lower the price each firm will charge. This makes sense: The more firms there are, the more competition each firm faces. The upward-sloping curve *CC* tells us that the more firms there are in the industry, the higher the average cost of each firm. This also makes sense: If the number of firms increases, each firm will sell less, so firms will not be able to move as far down their average cost curve.

The two schedules intersect at point E, corresponding to the number of firms n_2. The significance of n_2 is that it is the *zero-profit* number of firms in the industry. When there are n_2 firms in the industry, their profit-maximizing price is P_2, which is exactly equal to their average cost AC_2. What we will now argue is that in the long run, the number of firms in the industry tends to move toward n_2, so that point E describes the industry's long-run equilibrium.

To see why, suppose that n were less than n_2, say n_1. Then the price charged by firms would be P_1, while their average cost would be only AC_1. Thus firms would be making monopoly profits. Conversely, suppose that n were greater than n_2, say n_3. Then firms would charge only the price P_3, while their average cost would be AC_3. Firms would be suffering losses.

Over time, firms will enter an industry that is profitable and exit one in which they lose money. The number of firms will rise over time if it is less than n_2, fall if it is greater. This means that n_2 is the equilibrium number of firms in the industry and that P_2 is the equilibrium price.[5]

We have just developed a model of a monopolistically competitive industry in which we can determine the equilibrium number of firms and the average price that firms charge. We now use this model to derive some important conclusions about the role of economies of scale in international trade.

Monopolistic Competition and Trade

Underlying the application of the monopolistic competition model to trade is the idea that trade increases market size. In industries where there are economies of scale, both the variety of goods that a country can produce and the scale of its production are constrained by the size of the market. By trading with each other, and therefore forming an integrated world market that is bigger than any individual national market, nations are able to loosen these constraints. Each country can thus specialize in producing a narrower range of products than it would in the absence of trade; yet by buying from other countries the goods that it does not make, each nation can simultaneously increase the variety of goods available to its consumers. As a result, trade offers an opportunity for mutual gain even when countries do not differ in their resources or technology.

Suppose, for example, that there are two countries, each with an annual market for 1 million automobiles. By trading with each other, these countries can create a combined market of 2 million autos. In this combined market, more varieties of automobiles can be produced, at lower average costs, than in either market alone.

The monopolistic competition model can be used to show how trade improves the trade-off between scale and variety that individual nations face. We will begin by showing how a larger market leads, in the monopolistic competition model, to both a lower average price and the availability of a greater variety of goods. Applying this result to international trade, we observe that trade creates a world market larger than any of the national markets that comprise it. Integrating markets through international trade therefore has the same effects as growth of a market within a single country.

The Effects of Increased Market Size

The number of firms in a monopolistically competitive industry and the prices they charge are affected by the size of the market. In larger markets there usually will be both more firms and more sales per firm; consumers in a large market will be offered both lower prices and a greater variety of products than consumers in small markets.

---

[5]This analysis slips past a slight problem: The number of firms in an industry must, of course, be a whole number like 5 or 8. What if n_2 turns out to equal 6.37? The answer is that there will be six firms in the industry, all making small monopoly profits and not being challenged by new entrants because everyone knows that a seven-firm industry would lose money. In most examples of monopolistic competition, this whole-number or "integer constraint" problem turns out not to be very important, and we ignore it here.

To see this in the context of our model, look again at the *CC* curve in Figure 8-3, which showed that average costs per firm are higher the more firms there are in the industry. The definition of the *CC* curve is given by equation (8-6):

$$AC = F/Q + c = n \times F/S + c.$$

Examining this equation, we see that an increase in total industry output S will reduce average costs for any given number of firms n. The reason is that if the market grows while the number of firms is held constant, output per firm will increase and the average cost of each firm will therefore decline. Thus if we compare two markets, one with higher S than the other, the *CC* curve in the larger market will be below that in the smaller one.

Meanwhile, the *PP* curve in Figure 8-3, which relates the price charged by firms to the number of firms, does not shift. The definition of that curve was given in equation (8-10):

$$P = c + 1/(b \times n).$$

The size of the market does not enter into this equation, so an increase in S does not shift the *PP* curve.

Figure 8-4 uses this information to show the effect of an increase in the size of the market on long-run equilibrium. Initially, equilibrium is at point 1, with a price P_1 and a number of firms n_1. An increase in the size of the market, measured by industry sales S, shifts
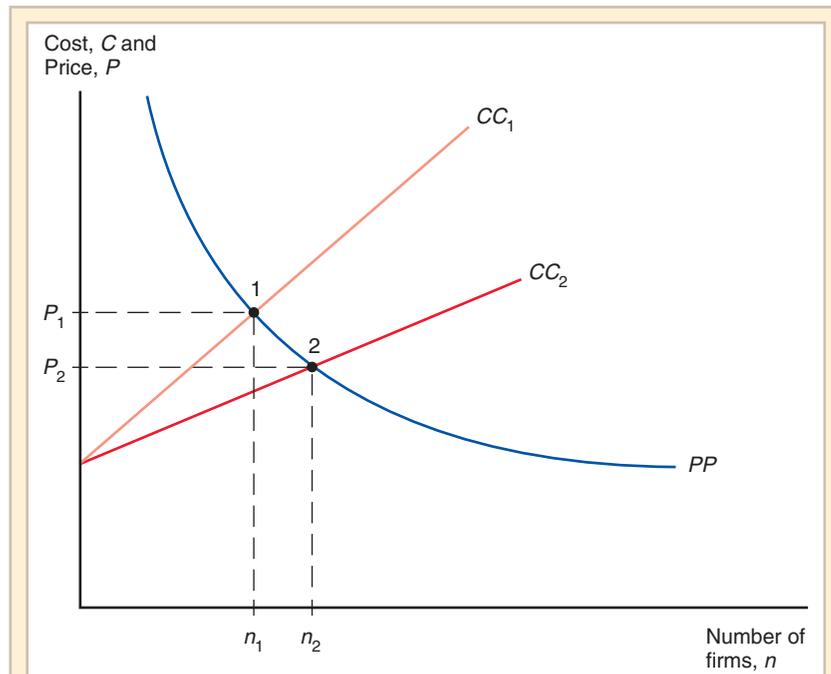
Figure 8-4

Effects of a Larger Market

An increase in the size of the market allows each firm, other things equal, to produce more and thus have lower average cost. This is represented by a downward shift from CC_1 to CC_2. The result is a simultaneous increase in the number of firms (and hence in the variety of goods available) and a fall in the price of each.

the *CC* curve down from CC_1 to CC_2, while it has no effect on the *PP* curve. The new equilibrium is at point 2: The number of firms increases from n_1 to n_2, while the price falls from P_1 to P_2.

Clearly, consumers would prefer to be part of a large market rather than a small one. At point 2, a greater variety of products is available at a lower price than at point 1.

Gains from an Integrated Market: A Numerical Example

International trade can create a larger market. We can illustrate the effects of trade on prices, scale, and the variety of goods available with a specific numerical example.

Imagine that automobiles are produced by a monopolistically competitive industry. The demand curve facing any given producer of automobiles is described by equation (8-5), with $b = 1/30,000$ (this value has no particular significance; it was chosen to make the example come out neatly). Thus the demand facing any one producer is given by

$$Q = S \times [(1/n) - (1/30,000) \times (P - \bar{P})],$$

where Q is the number of automobiles sold per firm, S is the total number sold for the industry, n is the number of firms, P is the price that a firm charges, and $\bar{P}$ is the average price of other firms. We also assume that the cost function for producing automobiles is described by equation (8-3), with a fixed cost $F = \$750,000,000$ and a marginal cost $c = \$5,000$ per automobile (again, these values were chosen to give nice results). The total cost is

$$C = 750,000,000 + (5,000 \times Q).$$

The average cost curve is therefore

$$AC = (750,000,000/Q) + 5,000.$$

Now suppose there are two countries, Home and Foreign. Home has annual sales of 900,000 automobiles; Foreign has annual sales of 1.6 million. The two countries are assumed, for the moment, to have the same costs of production.

Figure 8-5a shows the *PP* and *CC* curves for the Home auto industry. We find that in the absence of trade, Home would have six automobile firms, selling autos at a price of $10,000 each. (It is also possible to solve for n and P algebraically, as shown in the Mathematical Postscript to this chapter.) To confirm that this is the long-run equilibrium, we need to show both that the pricing equation (8-10) is satisfied and that the price equals average cost.

Substituting the actual values of the marginal cost c, the demand parameter b, and the number of Home firms n into equation (8-10), we find

$$P = \$10,000 = c + 1/(b \times n) = \$5,000 + 1/[(1/30,000) \times 6$$

$$= \$5,000 + \$5,000,$$

so the condition for profit maximization—marginal revenue equaling marginal cost—is satisfied. Each firm sells 900,000 units/6 firms = 150,000 units/firm. Its average cost is therefore

$$AC = (\$750,000,000/150,000) + \$5,000 = \$10,000.$$

Since the average cost of $10,000 per unit is the same as the price, all monopoly profits have been competed away. Thus six firms, selling each unit at a price of $10,000, with each firm producing 150,000 cars, is the long-run equilibrium in the Home market.
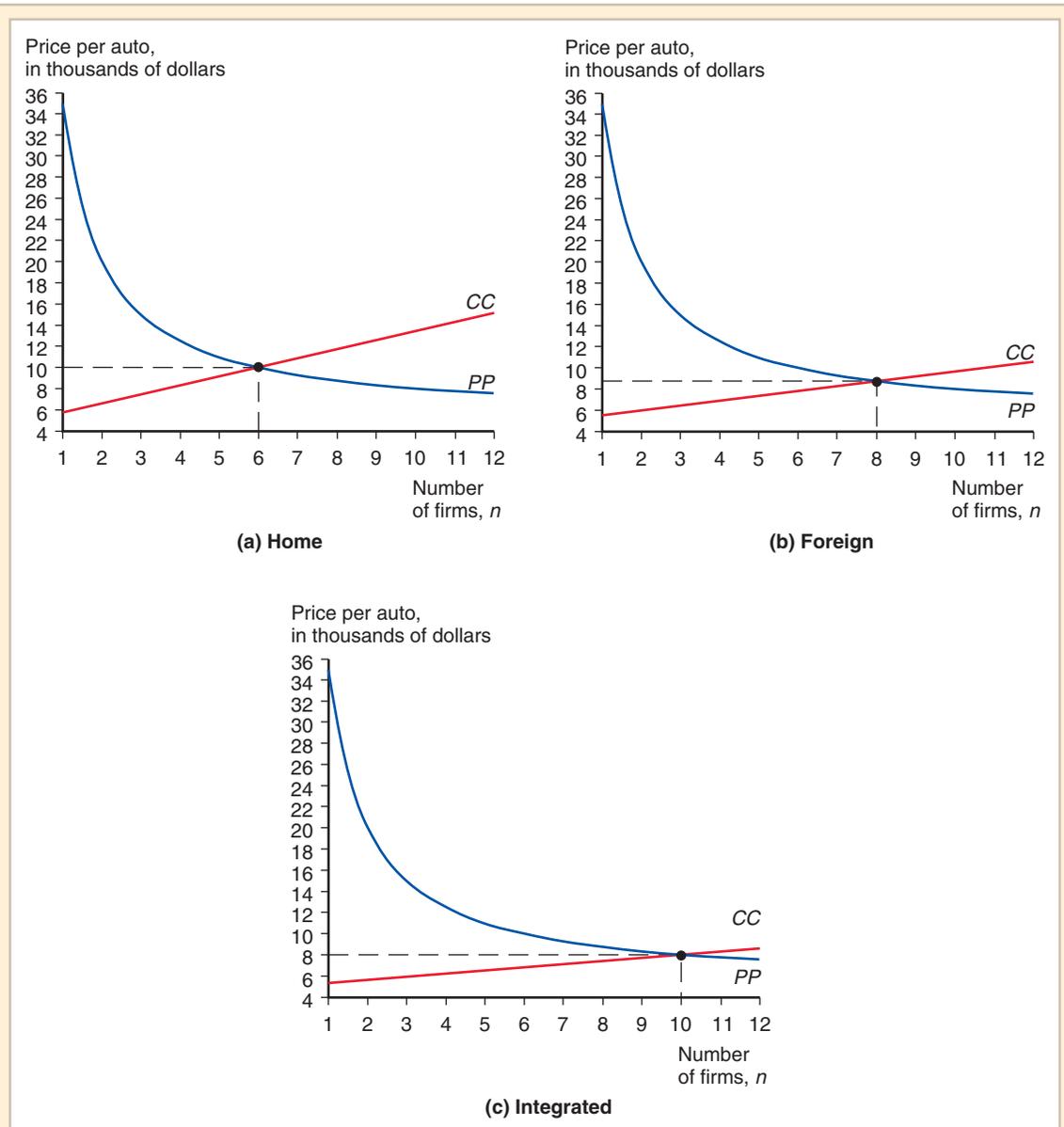
Figure 8-5

Equilibrium in the Automobile Market

(a) The Home market: With a market size of 900,000 automobiles, Home's equilibrium, determined by the intersection of the *PP* and *CC* curves, occurs with six firms and an industry price of $10,000 per auto. (b) The Foreign market: With a market size of 1.6 million automobiles, Foreign's equilibrium occurs with eight firms and an industry price of $8,750 per auto. (c) The combined market: Integrating the two markets creates a market for 2.5 million autos. This market supports ten firms, and the price of an auto is only $8,000.

What about Foreign? By drawing the *PP* and *CC* curves (panel (b) in Figure 8-5), we find that when the market is for 1.6 million automobiles, the curves intersect at $n = 8$, $P = 8{,}750$. That is, in the absence of trade, Foreign's market would support eight firms, each producing 200,000 automobiles, and selling them at a price of $8,750. We can again confirm that this solution satisfies the equilibrium conditions:

$$P = \$8{,}750 = c + 1/(b \times n) = \$5{,}000 + 1/[(1/30{,}000) \times 8] = \$5{,}000 + \$3{,}750,$$

and

$$AC = (\$750{,}000{,}000/200{,}000) + \$5{,}000 = \$8{,}750.$$

Now suppose it is possible for Home and Foreign to trade automobiles costlessly with one another. This creates a new, integrated market (panel (c) in Figure 8-5) with total sales of 2.5 million. By drawing the *PP* and *CC* curves one more time, we find that this integrated market will support ten firms, each producing 250,000 cars and selling them at a price of $8,000. The conditions for profit maximization and zero profits are again satisfied:

$$P = \$8{,}000 = c + 1/(b \times n) = \$5{,}000 + 1/[(1/30{,}000) \times 10]$$
$$= \$5{,}000 + \$3{,}000,$$

and

$$AC = (\$750{,}000{,}000/250{,}000) + \$5{,}000 = \$8{,}000.$$

We summarize the results of creating an integrated market in Table 8-1. The table compares each market alone with the integrated market. The integrated market supports more firms, each producing at a larger scale and selling at a lower price than either national market does on its own.

Clearly everyone is better off as a result of integration. In the larger market, consumers have a wider range of choices, yet each firm produces more and is therefore able to offer its product at a lower price. To realize these gains from integration, the countries must engage in international trade. To achieve economies of scale, each firm must concentrate its production in one country—either Home or Foreign. Yet it must sell its output to customers in both markets. So each product will be produced in only one country and exported to the other.

This numerical example highlights two important new features about trade with monopolistic competition relative to the models of trade based on comparative advantage that we covered in Chapters 3 through 6: (1) First, the example shows how product differentiation

TABLE 8-1	Hypothetical Example of Gains from Market Integration		
	Home Market, Before Trade	**Foreign Market, Before Trade**	**Integrated Market, After Trade**
Industry output (# of autos)	900,000	1,600,000	2,500,000
Number of firms	6	8	10
Output per firm (# of autos)	150,000	200,000	250,000
Average cost	$10,000	$8,750	$8,000
Price	$10,000	$8,750	$8,000

and internal economies of scale lead to trade between similar countries with no comparative advantage differences between them. This is a very different kind of trade than the one based on comparative advantage, where each country exports its comparative advantage good. Here, both Home and Foreign export autos to one another. Home pays for the imports of some automobile models (those produced by firms in Foreign) with exports of different types of models (those produced by firms in Home)—and vice versa. This leads to what is called **intra-industry trade:** two-way exchanges of similar goods. (2) Second, the example highlights two new channels for welfare benefits from trade. In the integrated market after trade, both Home and Foreign consumers benefit from a greater variety of automobile models (ten versus six or eight) at a lower price ($8,000 versus $8,750 or $10,000) as firms are able to consolidate their production destined for both locations and take advantage of economies of scale.[6]

Empirically, is intra-industry trade relevant and do we observe gains from trade in the form of greater product variety and consolidated production at lower average cost? The answer is yes.

The Significance of Intra-Industry Trade

The proportion of intra-industry trade in world trade has steadily grown over the last half-century. The measurement of intra-industry trade relies on an industrial classification system that categorizes goods into different industries. Depending on the coarseness of the industrial classification used (hundreds of different industry classifications versus thousands), intra-industry trade accounts for one-quarter to nearly one-half of all world trade flows. Intra-industry trade plays an even more prominent role in the trade of manufactured goods among advanced industrial nations, which accounts for the majority of world trade.

Table 8-2 shows measures of the importance of intra-industry trade for a number of U.S. manufacturing industries in 2009. The measure shown is intra-industry trade as a proportion of

TABLE 8-2	Indexes of Intra-Industry Trade for U.S. Industries, 2009
Metalworking Machinery	0.97
Inorganic Chemicals	0.97
Power-Generating Machines	0.86
Medical and Pharmaceutical Products	0.85
Scientific Equipment	0.84
Organic Chemicals	0.79
Iron and Steel	0.76
Road Vehicles	0.70
Office Machines	0.58
Telecommunications Equipment	0.46
Furniture	0.30
Clothing and Apparel	0.11
Footwear	0.10

[6]Also note that Home consumers gain more than Foreign consumers from trade integration. This is a standard feature of trade models with increasing returns and product differentiation: A smaller country stands to gain more from integration than a larger country. This is because the gains from integration are driven by the associated increase in market size; the country that is initially smaller benefits from a bigger increase in market size upon integration.

overall trade.[7] The measure ranges from 0.97 for metalworking machinery and inorganic chemicals—industries where U.S. exports and imports are nearly equal—to 0.10 for footwear, an industry in which the United States has large imports but virtually no exports. The measure would be 0 for an industry in which the United States is only an exporter or only an importer, but not both; it would be 1 for an industry in which U.S. exports exactly equal U.S. imports.

Table 8-2 shows that intra-industry trade is a very important component of trade for the United States in many different industries. Those industries tend to be ones that produce sophisticated manufactured goods, such as chemicals, pharmaceuticals, and specialized machinery. These goods are exported principally by advanced nations and are probably subject to important economies of scale in production. At the other end of the scale are the industries with very little intra-industry trade, which typically produce labor-intensive products such as footwear and apparel. These are goods that the United States imports primarily from less-developed countries, where comparative advantage is the primary determinant of U.S. trade with these countries.

What about the new types of welfare gains via increased product variety and economies of scale? A recent paper by Christian Broda at the Chicago Booth School of Business and David Weinstein at Columbia University estimates that the number of available products in U.S. imports tripled in the 30-year time-span from 1972 to 2001. They further estimate that this increased product variety for U.S. consumers represented a welfare gain equal to 2.6 percent of U.S. GDP![8]

Table 8-1 from our numerical example showed that the gains from integration generated by economies of scale were most pronounced for the smaller economy: Prior to integration, production there was particularly inefficient, as the economy could not take advantage of economies of scale in production due to the country's small size. This is exactly what happened when the United States and Canada followed a path of increasing economic integration starting with the North American Auto Pact in 1964 (which did not include Mexico) and culminating in the North American Free Trade Agreement (NAFTA, which does include Mexico). The Case Study that follows describes how this integration led to consolidation and efficiency gains in the automobile sector—particularly on the Canadian side (whose economy is one-tenth the size of the U.S. economy).

Similar gains from trade have also been measured for other real-world examples of closer economic integration. One of the most prominent examples has taken place in Europe over the last half-century. In 1957 the major countries of Western Europe established a free trade area in manufactured goods called the Common Market, or European Economic Community (EEC). (The United Kingdom entered the EEC later, in 1973.) The result was a rapid growth of trade that was dominated by intra-industry trade. Trade within the EEC grew twice as fast as world trade as a whole during the 1960s. This integration slowly expanded into what has become the European Union. When a subset of these countries (mostly, those countries that had formed the EEC) adopted the common euro currency in 1999, intra-industry trade among those countries further increased (even relative to that of the other countries in the European Union). Recent studies have also found that the adoption of the euro has led to a substantial increase in the number of different products that are traded within the Eurozone.

---

[7]To be more precise, the standard formula for calculating the importance of intra-industry trade within a given industry is

$$I = \frac{min\{\text{exports, imports}\}}{(\text{exports } + \text{ imports})/2},$$

where min{exports, imports} refers to the smallest value between exports and imports. This is the amount of two-way exchanges of goods that is reflected in *both* exports and imports. This number is measured as a proportion of the average trade flow (average of exports and imports). If trade in an industry flows in only one direction, then $I = 0$ since the smallest trade flow is zero: There is no intra-industry trade. On the other hand, if a country's exports and imports within an industry are equal, we get the opposite extreme of $I = 1$.

[8]See Christian Broda and David E. Weinstein, "Globalization and the Gains from Variety," *Quarterly Journal of Economics* 121 (April 2006), pp. 541–585.

Case Study

Intra-Industry Trade in Action: The North American Auto Pact of 1964

An unusually clear-cut example of the role of economies of scale in generating beneficial international trade is provided by the growth in automotive trade between the United States and Canada during the second half of the 1960s. While the case does not fit our model exactly since it involves multinational firms, it does show that the basic concepts we have developed are useful in the real world.

Before 1965, tariff protection by Canada and the United States produced a Canadian auto industry that was largely self-sufficient, neither importing nor exporting much. The Canadian industry was controlled by the same firms as the U.S. industry—a feature that we will address later on in this chapter—but these firms found it cheaper to have largely separate production systems than to pay the tariffs. Thus the Canadian industry was in effect a miniature version of the U.S. industry, at about $1/10$ the scale.

The Canadian subsidiaries of U.S. firms found that small scale was a substantial disadvantage. This was partly because Canadian plants had to be smaller than their U.S. counterparts. Perhaps more importantly, U.S. plants could often be "dedicated"—that is, devoted to producing a single model or component—while Canadian plants had to produce several different things, requiring the plants to shut down periodically to change over from producing one item to producing another, to hold larger inventories, to use less specialized machinery, and so on. The Canadian auto industry thus had a labor productivity about 30 percent lower than that of the United States.

The Ambassador bridge connects Detroit in the United States to Windsor in Canada. On a typical day, $250 million worth of cars and car parts crosses this bridge.

In an effort to remove these problems, the United States and Canada agreed in 1964 to establish a free trade area in automobiles (subject to certain restrictions). This allowed the auto companies to reorganize their production. Canadian subsidiaries of the auto firms sharply cut the number of products made in Canada. For example, General Motors cut in half the number of models assembled in Canada. The overall level of Canadian production and employment was, however, maintained. Production levels for the models produced in Canada rose dramatically, as those Canadian plants became one of the main (and many times the only) supplier of that model for the whole North American market. Conversely, Canada then imported the models from the United States that it was no longer producing. In 1962, Canada exported $16 million worth of automotive products to the United States while importing $519 million worth. By 1968 the numbers were $2.4 and $2.9 billion, respectively. In other words, both exports and imports increased sharply: intra-industry trade in action.

The gains seem to have been substantial. By the early 1970s the Canadian industry was comparable to the U.S. industry in productivity. Later on, this transformation of the automotive industry was extended to include Mexico. In 1989, Volkswagen consolidated its North American operations in Mexico, shutting down its plant in Pennsylvania. This process continued with the implementation of NAFTA (the North American Free Trade Agreement between the United States, Canada, and Mexico). In 1994 Volkswagen started producing the new Beetle for the whole North American market in that same Mexican plant. We discuss the effects of NAFTA in more detail later on in this chapter.

Firm Responses to Trade: Winners, Losers, and Industry Performance

In our numerical example of the auto industry with two countries, we saw how economic integration led to an increase in competition between firms. Of the 14 firms producing autos before trade (6 in Home and 8 in Foreign), only 10 firms "survive" after economic integration; however, each of those firms now produces at a bigger scale (250,000 autos produced per firm versus either 150,000 for Home firms or 200,000 for Foreign firms before trade). In that example, the firms were assumed to be symmetric, so exactly which firms exited and which survived and expanded was inconsequential. In the real world, however, performance varies widely across firms, so the effects of increased competition from trade are far from inconsequential. As one would expect, increased competition tends to hurt the worst-performing firms the hardest, because they are the ones who are forced to exit. If the increased competition comes from trade (or economic integration), then it is also associated with sales opportunities in new markets for the surviving firms. Again, as one would expect, it is the best-performing firms that take greatest advantage of those new sales opportunities and expand the most.

These composition changes have a crucial consequence at the level of the industry: When the better-performing firms expand and the worse-performing ones contract or exit, then overall industry performance improves. This means that trade and economic integration can have a direct impact on industry performance: It is as if there was technological growth at the level of the industry. Empirically, these composition changes generate substantial improvements in industry productivity.

Take the example of Canada's closer economic integration with the United States (see the preceding Case Study and the discussion in Chapter 2). We discussed how this integration led the automobile producers to consolidate production in a smaller number of Canadian plants, whose production levels rose dramatically. The Canada–U.S. Free Trade Agreement, which went into effect in 1989, extended the auto pact to most manufacturing sectors. A similar process of consolidation occurred throughout the affected Canadian manufacturing sectors. However, this was also associated with a selection process: The worst-performing producers shut down, while the better-performing ones expanded via large increases in exports to the U.S. market. Daniel Trefler at the University of Toronto has studied the effects of this trade agreement in great detail, examining the varied responses of Canadian firms.[9] He found that productivity in the most affected Canadian industries rose by a dramatic 14 to 15 percent (replicated economy-wide, a 1 percent increase in productivity translates into a 1 percent increase in GDP, holding employment constant). On its own, the contraction and exit of the worst-performing firms in response to increased competition from U.S. firms accounted for half of the 15 percent increase in those sectors.

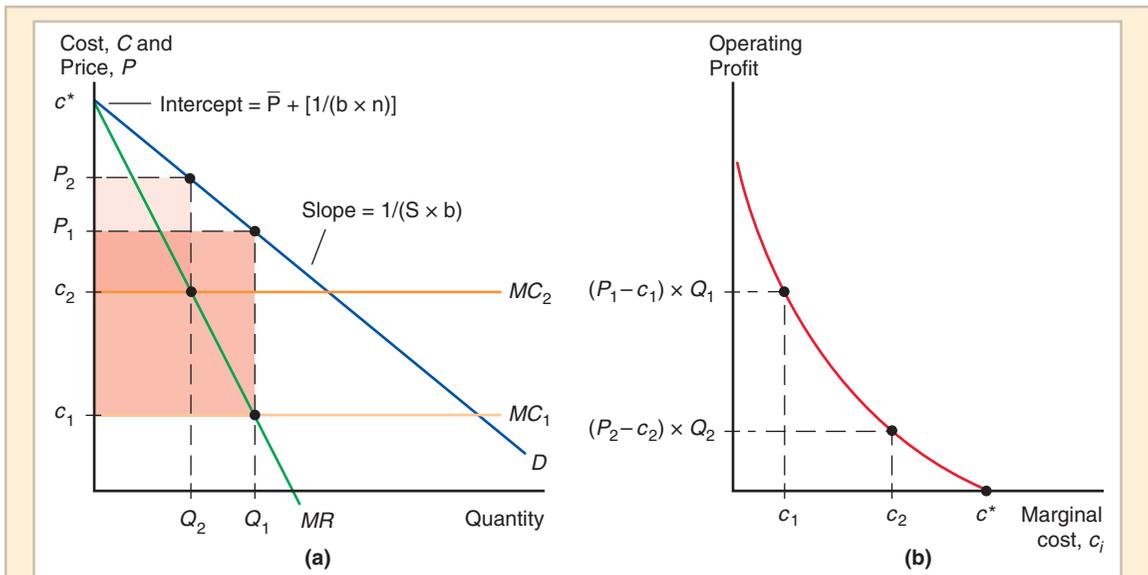## Performance Differences Across Producers

We now relax the symmetry assumption that we imposed in our previous development of the monopolistic competition model so that we can examine how competition from increased market size affects firms differently. The symmetry assumption meant that all firms had the same cost curve (8-3) and the same demand curve (8-5). Suppose now that

---

[9]See Daniel Trefler, "The Long and Short of the Canada-U.S. Free Trade Agreement," *American Economic Review* 94 (September 2004), pp. 870–895, and the summary of this work in the *New York Times:* "What Happened When Two Countries Liberalized Trade? Pain, Then Gain" by Virginia Postel (January 27, 2005).

firms have different cost curves because they produce with different marginal cost levels c_i. We assume that all firms still face the same demand curve. Product-quality differences between firms would lead to very similar predictions for firm performance as the ones we now derive for cost differences.

Figure 8-6 illustrates the performance differences between firms 1 and 2 when $c_1 < c_2$. In panel (a), we have drawn the common demand curve (8-5) as well as its associated marginal revenue curve (8-8). Note that both curves have the same intercept on the vertical axis (plug $Q = 0$ into (8-8) to obtain $MR = P$); this intercept is given by the price P from (8-5) when $Q = 0$, which is $\bar{P} + [1/(b \times n)]$. The slope of the demand curve is $1/(S \times b)$. As we previously discussed, the marginal revenue curve is steeper than the demand curve. Firms 1 and 2 choose output levels Q_1 and Q_2, respectively, to maximize their profits. This occurs where their respective marginal cost curves intersect the common marginal revenue curve. They set prices P_1 and P_2 that correspond to those output levels on the common demand curve. We immediately see that firm 1 will set a lower price and produce a higher output level than firm 2. Since the marginal revenue curve is steeper than the demand curve, we also see that firm 1 will set a higher markup over marginal cost than firm 2: $P_1 - c_1 > P_2 - c_2$.

The shaded areas represent operating profits for both firms, equal to revenue $P_i \times Q_i$ minus operating costs $c_i \times Q_i$ (for both firms, $i = 1$ and $i = 2$). Here, we have assumed that the fixed cost F (assumed to be the same for all firms) cannot be recovered and does not enter into operating profits (that is, it is a sunk cost). Since operating profits can be rewritten

Figure 8-6

Performance Differences Across Firms

(a) Demand and cost curves for firms 1 and 2. Firm 1 has a lower marginal cost than firm 2: $c_1 < c_2$. Both firms face the same demand curve and marginal revenue curve. Relative to firm 2, firm 1 sets a lower price and produces more output. The shaded areas represent operating profits for both firms (before the fixed cost is deducted). Firm 1 earns higher operating profits than firm 2. (b) Operating profits as a function of a firm's marginal cost c_i. Operating profits decrease as the marginal cost increases. Any firm with marginal cost above c^* cannot operate profitably and shuts down.

as the product of the markup times the number of output units sold,$(P_i - c_i) \times Q_i$, we can determine that firm 1 will earn higher profits than firm 2 (recall that firm 1 sets a higher markup and produces more output than firm 2). We can thus summarize all the relevant performance differences based on marginal cost differences across firms. Compared to a firm with a higher marginal cost, a firm with a lower marginal cost will: (1) set a lower price, but at a higher markup over marginal cost; (2) produce more output; and (3) earn higher profits.[10]

Panel (b) in Figure 8-6 shows how a firm's operating profits vary with its marginal cost c_i. As we just mentioned, this will be a decreasing function of marginal cost. Going back to panel (a), we see that a firm can earn a positive operating profit so long as its marginal cost is below the intercept of the demand curve on the vertical axis at $\bar{P} + [1/(b \times n)]$. Let c^* denote this cost cutoff. A firm with a marginal cost c_i above this cutoff is effectively "priced out" of the market and would earn negative operating profits if it were to produce any output. Such a firm would choose to shut down and not produce (incurring an overall profit loss equal to the fixed cost F). Why would such a firm enter in the first place? Clearly, it wouldn't if it knew about its high cost c_i prior to entering and paying the fixed cost F.

We assume that entrants face some randomness about their future production cost c_i. This randomness disappears only *after* F is paid and is sunk. Thus, some firms will regret their entry decision if their overall profit (operating profit minus the fixed cost F) is negative. On the other hand, some firms will discover that their production cost c_i is very low and that they earn high positive overall profit levels. Entry is driven by a similar process as the one we described for the case of symmetric firms. In that previous case, firms entered until profits for all firms were driven to zero. Here, there are profit differences between firms, and entry occurs until *expected* profits across all potential cost levels c_i are driven to zero.

## The Effects of Increased Market Size

Panel (b) of Figure 8-6 summarizes the industry equilibrium given a market size S. It tells us which range of firms survive and produce (with cost c_i below c^*), and how their profits will vary with their cost levels c_i. What happens when economies integrate into a single larger market? As was the case with symmetric firms, a larger market can support a larger number of firms than can a smaller market. This also implies more competition in the larger market. What are the repercussions for different firms of increased competition?

First, consider the effects of increased competition (higher number of firms n) on the individual firm-demand curves. Panel (a) of Figure 8-7 shows the effect. Recall that the intercept on the vertical axis is equal to $\bar{P} + [1/(b \times n)]$, which decreases when the number of firms increases.[11] The slope of the demand curve, equal to $1/(S \times b)$, decreases from the direct effect of the increase in the market size S, so the demand curve also becomes flatter: With increased competition, a producer can gain more market share from a given price cut. This produces the shift in the demand curve from D to D' shown in panel (a) of Figure 8-7. Notice how the demand curve shifts in for the smaller firms (lower-output Q_i) that operate on the top part of the demand curve.

Panel (b) of Figure 8-7 shows the consequences of this demand change for the operating profits of firms with different cost levels c_i. The decrease in demand for the smaller firms translates into a new, lower-cost cutoff, $c^{*\prime}$: Some firms with the high cost levels above $c^{*\prime}$ cannot survive the decrease in demand and are forced to exit. On the other hand,

[10]Recall that we have assumed that all firms face the same nonrecoverable fixed cost F. If a firm earns higher operating profits, then it also earns higher overall profits (that deduct the fixed cost F).
[11]The intercept will further decrease because the average price will also decrease.
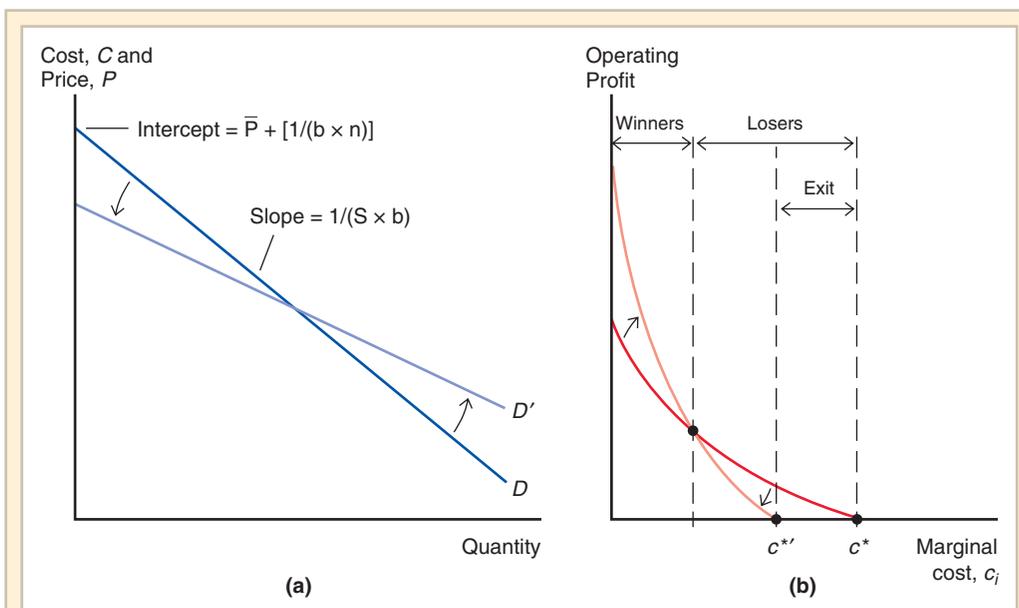
Figure 8-7

Winners and Losers from Economic Integration

(a) The demand curve for all firms shifts from D to D'. It is flatter, and has a lower intercept on the vertical axis. (b) Effects of the shift in demand on the operating profits of firms with different marginal cost c_i. Firms with marginal cost between the old cutoff, c^*, and the new one, $c^{*'}$, are forced to exit. Some firms with the lowest marginal cost levels gain from integration and their profits increase.

the flatter demand curve is advantageous to some firms with low cost levels: They can adapt to the increased competition by lowering their markup (and hence their price) and gain some additional market share.[12] This translates into increased profits for some of the best-performing firms with the lowest cost levels c_i.[13]

Figure 8-7 illustrates how increased market size generates both winners and losers among firms in an industry. The low-cost firms thrive and increase their profits and market shares, while the high-cost firms contract and the highest-cost firms exit. These composition changes imply that overall productivity in the industry is increasing as production is concentrated among the more productive (low-cost) firms. This replicates the findings for Canadian manufacturing following closer integration with U.S. manufacturing, as we previously described. These effects tend to be most pronounced for smaller countries that integrate with larger ones, but it is not limited to those small countries. Even for a big economy such as the United States, increased integration via lower trade costs leads to important composition effects and productivity gains.[14]

---

[12]Recall that the lower the firm's marginal cost c_i, the higher its markup over marginal cost $P_i - c_i$. High-cost firms are already setting low markups and cannot lower their prices to induce positive demand, as this would mean pricing below their marginal cost of production.

[13]Another way to deduce that profit increases for some firms is to use the entry condition that drives average profits to zero: If profit decreases for some of the high-cost firms, then it must increase for some of the low-cost firms, since the average across all firms must remain equal to zero.

[14]See A. B. Bernard, J. B. Jensen, and P. K. Schott, "Trade Costs, Firms and Productivity," *Journal of Monetary Economics* 53 (July 2006), pp. 917–937.

Trade Costs and Export Decisions

Up to now, we have modeled economic integration as an increase in market size. This implicitly assumes that this integration occurs to such an extent that a single combined market is formed. In reality, integration rarely goes that far: Trade costs among countries are reduced, but they do not disappear. In Chapter 2, we discussed how these trade costs are manifested even for the case of the two very closely integrated economies of the United States and Canada. We saw how the U.S.–Canada border substantially decreases trade volumes between Canadian provinces and U.S. states.

Trade costs associated with this border crossing are also a salient feature of firm-level trade patterns: Very few firms in the United States reach Canadian customers. In fact, most U.S. firms do not report *any* exporting activity at all (because they sell only to U.S. customers). In 2002, only 18 percent of U.S. manufacturing firms reported undertaking some export sales. Table 8-3 shows the proportion of firms that report some export sales across several different U.S. manufacturing sectors. Even in industries where exports represent a substantial proportion of total production, such as chemicals, machinery, electronics, and transportation, fewer than 40 percent of firms export. In fact, one major reason why trade costs associated with national borders reduce trade so much is that they drastically cut down the number of firms willing or able to reach customers across the border. (The other reason is that the trade costs also reduce the export sales of firms that do reach those customers across the border.)

In our integrated economy without any trade costs, firms were indifferent as to the location of their customers. We now introduce trade costs to explain why firms actually do care about the location of their customers, and why so many firms choose not to reach customers in another country. As we will see shortly, this will also allow us to explain important differences between those firms that choose to incur the trade costs and export, and those that do not. Why would some firms choose not to export? Simply put, the trade costs reduce the profitability of exporting for all firms. For some, that reduction in profitability makes exporting unprofitable. We now formalize this argument.

To keep things simple, we will consider the response of firms in a world with two identical countries (Home and Foreign). Let the market size parameter S now reflect the size of each market, so that $2 \times S$ now reflects the size of the world market. We cannot analyze this world market as a single market of size $2 \times S$ because this market is no longer perfectly integrated due to trade costs.

| TABLE 8-3	Proportion of U.S. Firms Reporting Export Sales by Industry, 2002
Printing	5%
Furniture	7%
Apparel	8%
Wood Products	8%
Fabricated Metals	14%
Petroleum and Coal	18%
Transportation Equipment	28%
Machinery	33%
Chemicals	36%
Computer and Electronics	38%
Electrical Equipment and Appliances	38%

Source: A. B. Bernard, J. B. Jensen, S. J. Redding, and P. K. Schott, "Firms in International Trade," *Journal of Economic Perspectives* 21 (Summer 2007), pp. 105–130.

Specifically, assume that a firm must incur an additional cost t for each unit of output that it sells to customers across the border. We now have to keep track of the firms' behavior in each market separately. Due to the trade cost t, firms will set different prices in their export market relative to their domestic market. This will lead to different quantities sold in each market, and ultimately to different profit levels earned in each market. As each firm's marginal cost is constant (does not vary with production levels), those decisions regarding pricing and quantity sold in each market can be separated: A decision regarding the domestic market will have no impact on the profitability of different decisions for the export market.

Consider the case of firms located in Home. Their situation regarding their domestic (Home) market is exactly as was illustrated in Figure 8-6, except that all the outcomes, such as price, output, and profit, relate to the domestic market only.[15] Now consider the decisions of firms 1 and 2 (with marginal costs c_1 and c_2) regarding the export (Foreign) market. They face the same demand curve in Foreign as they do in Home (recall that we assumed that the two countries are identical). The only difference is that the firms' marginal cost in the export market is shifted up by the trade cost t. Figure 8-8 shows the situation for the two firms in both markets.

What are the effects of the trade cost on the firms' decisions regarding the export market? We know from our previous analysis that a higher marginal cost induces a firm to raise its price, which leads to a lower output quantity sold and lower profits. We also know that if marginal cost is raised above a threshold level c^*, then a firm cannot profitably operate in that market. This is what happens to firm 2 in Figure 8-8. Firm 2 can profitably operate in
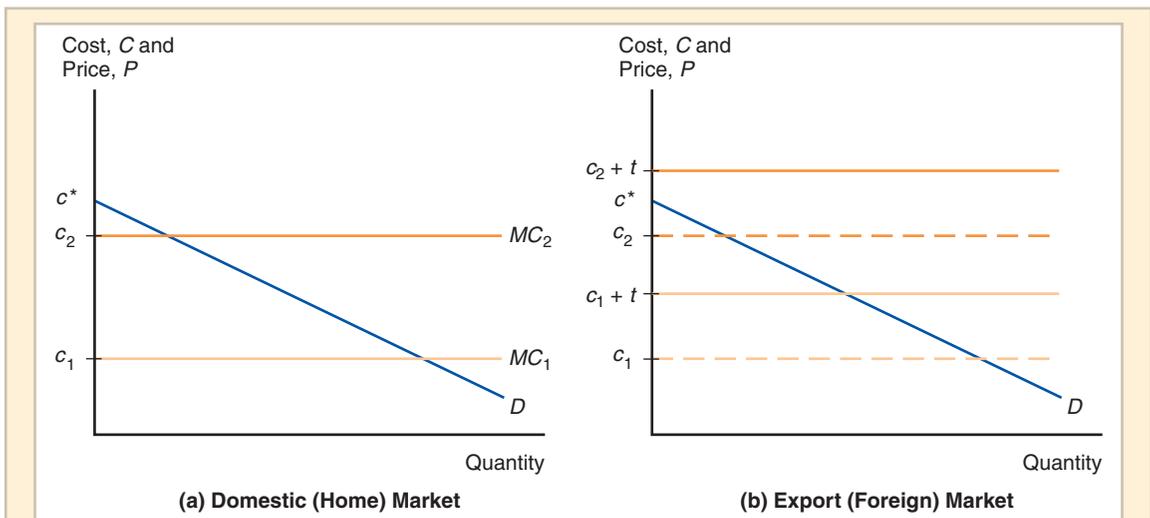
Figure 8-8

Export Decisions with Trade Costs

(a) Firms 1 and 2 both operate in their domestic (Home) market. (b) Only firm 1 chooses to export to the Foreign market. It is not profitable for firm 2 to export given the trade cost t.

---

[15]The number of firms n is the total number of firms selling in the Home market. (This includes both firms located in Home as well as the firms located in Foreign that export to Home). $\overline{P}$ is the average price across all those firms selling in Home.

its domestic market, because its cost there is below the threshold: $c_2 \leq c^*$. However, it cannot profitably operate in the export market because its cost there is above the threshold: $c_2 + t > c^*$. Firm 1, on the other hand, has a low enough cost that it can profitably operate in both the domestic and the export markets: $c_1 + t \leq c^*$. We can extend this prediction to all firms based on their marginal cost c_i. The lowest-cost firms with $c_i \leq c^* - t$ export; the higher-cost firms with $c^* - t < c_i \leq c^*$ still produce for their domestic market but do not export; the highest-cost firms with $c_i > c^*$ cannot profitably operate in either market, and thus exit.

We just saw how the modeling of trade costs added two important predictions to our model of monopolistic competition and trade: Those costs explain why only a subset of firms export, and they also explain why this subset of firms will consist of relatively larger and more productive firms (those firms with lower marginal cost c_i). Empirical analyses of firms' export decisions from numerous countries have provided overwhelming support for this prediction that exporting firms are bigger and more productive than firms in the same industry that do not export. In the United States in a typical manufacturing industry, an exporting firm is on average more than twice as large as a firm that does not export. The average exporting firm also produces 11 percent more value added (output minus intermediate inputs) per worker than the average nonexporting firm. These differences across exporters and nonexporters are even larger in many European countries.[16]

Dumping

Adding trade costs to our model of monopolistic competition also added another dimension of realism: Because markets are no longer perfectly integrated through costless trade, firms can choose to set different prices in different markets. The trade costs also affect how a firm responds to competition in a market. Recall that a firm with a higher marginal cost will choose to set a lower markup over marginal cost (this firm faces more intense competition due to its lower market share). This means that an exporting firm will respond to the trade cost by lowering its markup for the export market.

Consider the case of firm 1 in Figure 8-8. It faces a higher marginal cost $c_1 + t$ in the Foreign export market. Let P_1^D and P_1^X denote the prices that firm 1 sets on its domestic (Home) market and export (Foreign) market, respectively. Firm 1 sets a lower markup $P_1^X - (c_1 + t)$ on the export market relative to its markup $P_1^D - c_1$ on the domestic market. This in turn implies that $P_1^X - t < P_1^D$, and that firm 1 sets an export price (net of trade costs) that is lower than its domestic price.

That is considered **dumping** by firm 1, and is regarded by most countries as an "unfair" trade practice. Any firm from Foreign can appeal to its local authorities (in the United States, the Commerce Department and the International Trade Commission are the relevant authorities) and seek punitive damages against firm 1. This usually takes the form of an **antidumping duty** imposed on firm 1, and would usually be scaled to the price difference between P_1^D and $P_1^X - t$.[17]

---

[16]See A. B. Bernard, J. B. Jensen, S. J. Redding, and P. K. Schott, "Firms in International Trade," *Journal of Economic Perspectives* 21 (Summer 2007), pp. 105–130; and Thierry Mayer and Gianmarco I. P. Ottaviano, "The Happy Few: The Internationalisation of European Firms: New Facts Based on Firm-Level Evidence," *Intereconomics* 43 (May/June 2008), pp. 135–148.

[17]$P_1^X - t$ is called firm 1's *ex* factory price for the export market (the price at the "factory gate" before the trade costs are incurred). If firm 1 incurred some transport or delivery cost in its domestic market, then those costs would be deducted from its domestic price P_1^D to obtain an *ex factory* price for the domestic market. Antidumping duties are based on differences between a firm's ex factory prices in the domestic and export markets.

Dumping is a controversial issue in trade policy; we discuss policy disputes surrounding dumping in Chapter 10. For now, we just note that firm 1 is not behaving any differently than the foreign firms it is competing against in the Foreign market. In that market, firm 1 sets exactly the same markup over marginal cost as Foreign firm 2 with marginal cost $c_2 = c_1 + t$. Firm 2's pricing behavior is perfectly legal, so why is firm 1's export pricing decision considered to represent an "unfair" trade practice? This is one major reason why economists believe that the enforcement of dumping claims is misguided (see the Case Study below for other reasons) and that there is no good economic justification for dumping to be considered particularly harmful.

Our model of monopolistic competition highlighted how trade costs have a natural tendency to induce firms to lower their markups in export markets, where they face more intense competition due to their reduced market share. This makes it relatively easy for domestic firms to file a dumping complaint against exporters in their markets. In practice, those antidumping laws can then be used to erect barriers to trade by discriminating against exporters in a market.

Case Study

Antidumping as Protectionism

In the United States and a number of other countries, dumping is regarded as an unfair competitive practice. U.S. firms that claim to have been injured by foreign firms that dump their products in the domestic market at low prices can appeal, through a quasi-judicial procedure, to the Commerce Department for relief. If their complaint is ruled valid, an "antidumping duty" is imposed, equal to the calculated difference between the actual and the "fair" price of imports. In practice, the Commerce Department accepts the great majority of complaints by U.S. firms about unfair foreign pricing. The determination that this unfair pricing has actually caused injury, however, is in the hands of a different agency, the International Trade Commission, which rejects about half of its cases.

Economists have never been very happy with the idea of singling out dumping as a prohibited practice. For one thing, setting different prices for different customers is a perfectly legitimate business strategy—like the discounts that airlines offer to students, senior citizens, and travelers who are willing to stay over a weekend. Also, the legal definition of dumping deviates substantially from the economic definition. Since it is often difficult to prove that foreign firms charge higher prices to domestic than to export customers, the United States and other nations instead often try to calculate a supposedly fair price based on estimates of foreign production costs. This "fair price" rule can interfere with perfectly normal business practices: A firm may well be willing to sell a product for a loss while it is lowering its costs through experience or breaking into a new market.

In spite of almost universally negative assessments from economists, however, formal complaints about dumping have been filed with growing frequency since about 1970. China has attracted a particularly large number of antidumping suits, for two reasons. One is that China's rapid export growth has raised many complaints. The other is the fact that China is still nominally a communist country, and the United States officially considers it a

"nonmarket economy." A *Business Week* story described the difference that China's status makes: "That means the U.S. can simply ignore Chinese data on costs on the assumption they are distorted by subsidized loans, rigged markets, and the controlled yuan. Instead, the government uses data from other developing nations regarded as market economies. In the TV and furniture cases, the U.S. used India—even though it is not a big exporter of these goods. Since India's production costs were higher, China was ruled guilty of dumping."[18]

As the quote suggests, China has been subject to antidumping duties on TVs and furniture, along with a number of other products including crepe paper, hand trucks, shrimp, ironing tables, plastic shopping bags, steel fence posts, iron pipe fittings, and saccharin. These duties are high: as high as 78 percent on color TVs and 330 percent on saccharin.

Multinationals and Outsourcing

When is a corporation multinational? In U.S. statistics, a U.S. company is considered foreign-controlled, and therefore a subsidiary of a foreign-based multinational, if 10 percent or more of its stock is held by a foreign company; the idea is that 10 percent is enough to convey effective control. Similarly, a U.S.-based company is considered multinational if it owns more than 10 percent of a foreign firm. The controlling (owning) firm is called the multinational parent, while the "controlled" firms are called the multinational affiliates.

When a U.S. firm buys more than 10 percent of a foreign firm, or when a U.S. firm builds a new production facility abroad, that investment is considered a U.S. outflow of **foreign direct investment (FDI)**. The latter is called *greenfield* FDI, while the former is called *brownfield* FDI (or cross-border mergers and acquisitions). Conversely, investments by foreign firms in production facilities in the United States are considered U.S. FDI inflows. We describe the worldwide patterns of FDI flows in the Case Study that follows. For now, we focus on the decision of a firm to become a multinational parent. Why would a firm choose to operate an affiliate in a foreign location?

Case Study

Patterns of Foreign Direct Investment Flows Around the World

Figure 8-9 shows how the magnitude of worldwide FDI flows has evolved over the last 30 years. We first examine patterns for the world, where FDI flows must be balanced: Hence world inflows are equal to world outflows. We see that there was a massive increase in multinational activity in the mid- to late 1990s, when worldwide FDI flows more than quintupled, and then again in the early 2000s. We also see that the growth rate of FDI is very uneven, with huge peaks and troughs. Those peaks and troughs correlate with the gyrations of stock markets worldwide (strongly dominated by fluctuations in the U.S. stock market). The financial collapse in 2000 (the bursting of the dotcom bubble) and the most recent financial crisis in 2007–2009 also induced huge crashes in worldwide FDI flows. Most of those FDI flows related to cross-border mergers and acquisitions, whereas greenfield FDI remained relatively stable.

---

[18]"Wielding a Heavy Weapon Against China," *Business Week*, June 21, 2004.
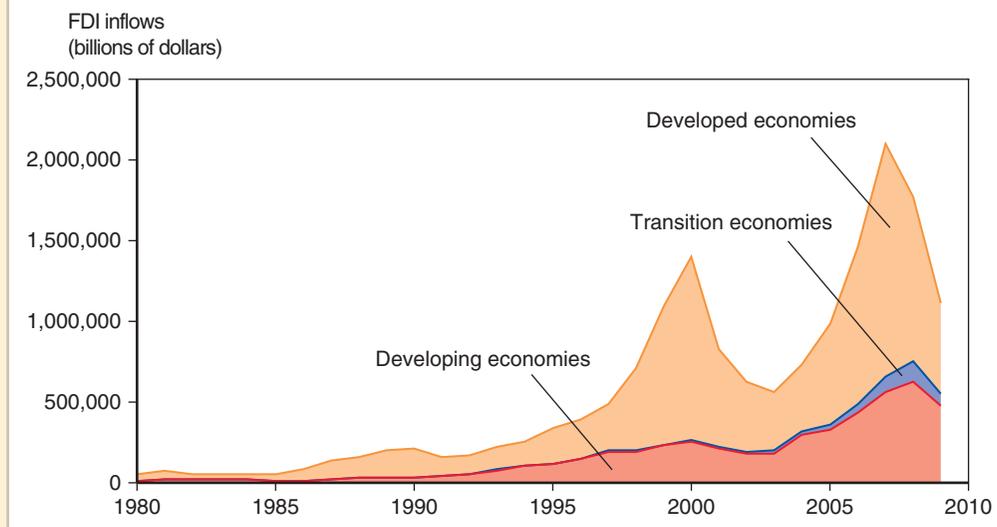
Figure 8-9

**Inflows of Foreign Direct Investment, 1980–2009 (billions of dollars)**

Worldwide flows of FDI have significantly increased since the mid-1990s, though the rates of increase have been very uneven. Historically, most of the inflows of FDI have gone to developed countries. However, the proportion of FDI inflows going to developing and transition economies has steadily increased over time and accounted for half of worldwide FDI flows in 2009.

Source: UNCTAD, World Investment Report, 2010.

Looking at the distribution of FDI inflows across groups of countries, we see that historically, developed countries have been the biggest recipients of inward FDI. However, we also see that those inflows are much more volatile (this is where the FDI related to mergers and acquisitions is concentrated) than the FDI going to developing and transition economies (economies in Central/Eastern Europe that used to be part of the Soviet Union or Yugoslavia). Finally, we can see that there has been a steady expansion in the share of FDI that flows to developing and transition countries. This accounted for half of worldwide FDI flows in 2009, after the most recent contraction in the flows to developed economies.

Figure 8-10 shows the list of the top 25 countries whose firms engage in FDI outflows. Because those flows are very volatile, especially with the recent crisis, they have been averaged over the past three years. We see that FDI outflows are still dominated by the developed economies; but we also see that big developing countries, most notably China (including Hong Kong), are playing an increasingly important role. In fact, one of the fastest-growing FDI segments is flows *from* developing countries *into* other developing countries. Multinationals in both China and India play a prominent role in this relatively new type of FDI. We also see that international tax policies can shape the location of FDI. For example, the British Virgin Islands would not figure in that top-25 list were it not for its status as an international tax haven. Firms from that location that engage in FDI are mainly offshore companies: They are incorporated in the British Virgin Islands, but their productive activities are located elsewhere in the world.

FDI flows are not the only way to measure the presence of multinationals in the world economy. Other measures are based on economic activities such as sales, value
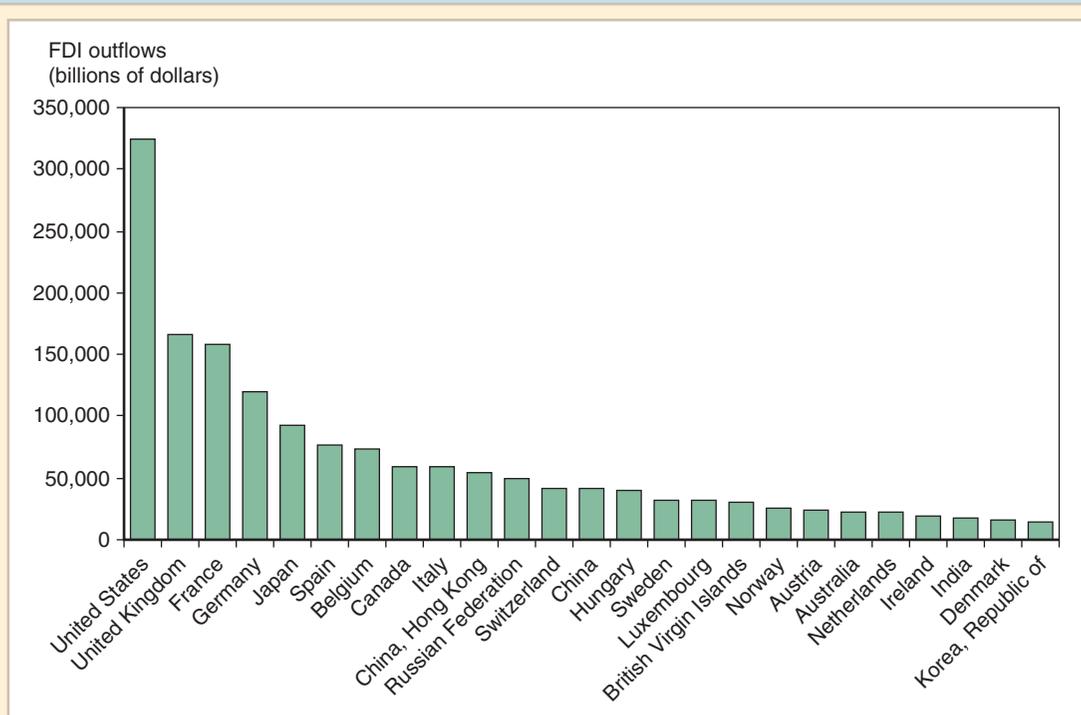
Figure 8-10

Outward Foreign Direct Investment for Top 25 Countries, Yearly Average for 2007–2009 (billions of dollars)

Developed countries dominate the list of the top countries whose firms engage in outward FDI. More recently, firms from some big developing countries such as China and India have performed significantly more FDI.

Source: UNCTAD, World Investment Report, 2010.

added (sales minus purchased intermediate goods), and employment. Sales of FDI affiliates are often used as the benchmark of multinational activity. This provides the relevant benchmark when comparing the activities of multinationals to export volumes. However, the sales of multinationals are also often compared to country GDPs showing, for example, that the big multinationals have higher sales volumes than the GDPs of many countries in the world. For the world as a whole in 2000, the total sales of the largest multinationals (top 200) amounted to more than 27 percent of world GDP.

However striking, this comparison is misleading and overstates the influence of multinationals, because country GDP is measured in terms of value added: Intermediate goods used in final production are not double-counted in this GDP measure. On the other hand, the intermediate goods that one multinational sells to another are double-counted in the multinationals' sales totals (once in the sales of the producer of the intermediate goods, and another time as part of the final value of the goods sold by the user of the intermediate goods). As a result, the appropriate comparison between multinationals and GDPs should be based on value added. By this metric, the value added produced by the biggest multinationals accounted for 4.3 percent of world GDP in 2000. This is still a big percentage, but not as eye-catching as the 27 percent measure.

The answer depends, in part, on the production activities that the affiliate carries out. These activities fall into two main categories: (1) The affiliate replicates the production process (that the parent firm undertakes in its domestic facilities) elsewhere in the world; and (2) the production chain is broken up, and parts of the production processes are transferred to the affiliate location. Investing in affiliates that do the first type of activities is categorized as **horizontal FDI**. Investing in affiliates that do the second type of activities is categorized as **vertical FDI**.[19]

Vertical FDI is mainly driven by production cost differences between countries (for those parts of the production process that can be performed in another location). What drives those cost differences between countries? This is just the outcome of the theory of comparative advantage that we developed in Chapters 3 through 7. For example, Intel (the world's largest computer chip manufacturer) has broken up the production of chips into wafer fabrication, assembly, and testing. Wafer fabrication and the associated research and development are very skill-intensive, so Intel still performs most of those activities in the United States, as well as in Ireland and Israel (where skilled labor is still relatively abundant). On the other hand, chip assembly and testing are labor-intensive, and Intel has moved those production processes to countries where labor is relatively abundant, such as Malaysia, the Philippines, and, more recently, Costa Rica and China. This type of vertical FDI is one of the fastest-growing types of FDI, and is behind the large increase in FDI inflows to developing countries (see Figure 8-9).

In contrast to vertical FDI, horizontal FDI is dominated by flows between developed countries; that is, both the multinational parent and the affiliates are located in developed countries. The main reason for this type of FDI is to locate production near a firm's large customer bases. Hence, trade and transport costs play a much more important role than production cost differences for these FDI decisions. Consider the example of Toyota, which is the world's largest motor vehicle producer (at least, at the time of writing). At the start of the 1980s, Toyota produced almost all of its cars and trucks in Japan and exported them throughout the world, but mostly to North America and Europe. High trade costs to those markets (in large part due to trade restrictions; see Chapter 9) and rising demand levels there induced Toyota to slowly expand its production overseas. By 2009, Toyota produced over half of its vehicles in assembly plants abroad. Toyota has replicated the production process for its most popular car model, the Corolla, in assembly plants in Japan, Canada, the United States, the United Kingdom, and Turkey: This is horizontal FDI in action.

The Firm's Decision Regarding Foreign Direct Investment

We now examine in more detail the firm's decision regarding horizontal FDI. We mentioned that one main driver was high trade costs associated with exporting, which leads to an incentive to locate production near customers. On the other hand, there are also increasing returns to scale in production. As a result, it is not cost effective to replicate the production process too many times and operate facilities that produce too little output to take advantage of those increasing returns. This is called the *proximity-concentration* trade-off for FDI. Empirical evidence on the extent of FDI across sectors strongly confirms the relevance of this trade-off: FDI activity is concentrated in sectors where trade costs are high (such as the automobile industry); however, when increasing returns to scale are important and average plant sizes are large, one observes higher export volumes relative to FDI.

[19]In reality, the distinctions between horizontal and vertical FDI can be blurred. Some large multinational parents operate large networks of affiliates that replicate parts of the production process, but are also vertically connected to other affiliates in the parent's network. This is referred to as "complex" FDI.

Empirical evidence also shows that there is an even stronger sorting pattern for FDI at the firm level *within* industries: Multinationals tend to be substantially larger and more productive than nonmultinationals in the same country. Even when one compares multinationals to the subset of exporting firms in a country, one still finds a large size and productivity differential in favor of the multinationals. We return to our monopolistic competition model of trade to analyze how firms respond differently to the proximity-concentration trade-off involved with the FDI decision.

The Horizontal FDI Decision   How does the proximity trade-off fit into our model of firms' export decisions captured in Figure 8-8? There, if a firm wants to reach customers in Foreign, it has only one possibility: export, and incur the trade cost t per unit exported. Let's now introduce the choice of becoming a multinational via horizontal FDI: A firm could avoid the trade cost t by building a production facility in Foreign. Of course, building this production facility is costly, and implies incurring the fixed cost F again for the foreign affiliate. (Note, however, that this additional fixed cost need not equal the fixed cost of building the firm's original production facility in Home; characteristics that are specific to the individual country will affect this cost.) For simplicity, continue to assume that Home and Foreign are similar countries so that this firm could build a unit of a good at the same marginal cost in this foreign facility. (Recall that horizontal FDI mostly involves developed countries with similar factor prices.)

The firm's export versus FDI choice will then involve a trade-off between the per-unit export cost t and the fixed cost F of setting up an additional production facility. Any such trade-off between a per-unit and a fixed cost boils down to scale. If the firm sells Q units in the foreign market, then it incurs a total trade-related cost $Q \times t$ to export; this is weighed against the alternative of the fixed cost F. If $Q > F/t$, then exporting is more expensive, and FDI is the profit-maximizing choice.

This leads to a scale cutoff for FDI. This cutoff summarizes the proximity-concentration trade-off: Higher trade costs on one hand, and lower fixed production costs on the other hand, both lower the FDI cutoff. The firm's scale, however, depends on its performance measure. A firm with low enough cost c_i will want to sell more than Q units to foreign customers. The most cost-effective way to do this is to build an affiliate in Foreign and become a multinational. Some firms with intermediate cost levels will still want to serve customers in Foreign, but their intended sales Q are low enough that exports, rather than FDI, will be the most cost-effective way to reach those customers.

The Vertical FDI Decision   A firm's decision to break up its production chain and move parts of that chain to a foreign affiliate will also involve a trade-off between per-unit and fixed costs—so the scale of the firm's activity will again be a crucial element determining this outcome. When it comes to vertical FDI, the key cost saving is not related to the shipment of goods across borders; rather, it involves production cost differences for the parts of the production chain that are being moved. As we previously discussed, those cost differences stem mostly from comparative advantage forces.

We will not discuss those cost differences further here, but rather ask why—given those cost differences—all firms do not choose to operate affiliates in low-wage countries to perform the activities that are most labor-intensive and can be performed in a different location. The reason is that, as with the case of horizontal FDI, vertical FDI requires a substantial fixed cost investment in a foreign affiliate in a country with the appropriate characteristics.[20]

---

[20]Clearly, factor prices such as wages are a crucial component, but other country characteristics, such as its transportation/public infrastructure, the quality of its legal institutions, and its tax/regulation policies toward multinationals, can be critical as well.

Again, as with the case of horizontal FDI, there will be a scale cutoff for vertical FDI that depends on the production cost differentials on one hand, and the fixed cost of operating a foreign affiliate on the other hand. Only those firms operating at a scale above that cutoff will choose to perform vertical FDI.

Outsourcing

Our discussion of multinationals up to this point has neglected an important motive. We discussed the **location motive** for production facilities that leads to multinational formation. However, we did not discuss why the parent firm chooses to *own* the affiliate in that location and operate as a single multinational firm. This is known as the **internalization motive**.

As a substitute for horizontal FDI, a parent could license an independent firm to produce and sell its products in a foreign location; as a substitute for vertical FDI, a parent could contract with an independent firm to perform specific parts of the production process in the foreign location with the best cost advantage. This substitute for vertical FDI is known as **foreign outsourcing** (sometimes just referred to as outsourcing, where the foreign location is implied).

Offshoring represents the relocation of parts of the production chain abroad and groups together both foreign outsourcing and vertical FDI. Offshoring has increased dramatically in the last decade and is one of the major drivers of the increased worldwide trade in services (such as business and telecommunications services); in manufacturing, trade in intermediate goods accounted for 40 percent of worldwide trade in 2008. When the intermediate goods are produced within a multinational's affiliate network, the shipments of those intermediate goods are classified as intra-firm trade. Intra-firm trade represents roughly one-third of worldwide trade and over 40 percent of U.S. trade.

What are the key elements that determine this internalization choice? Control over a firm's proprietary technology offers one clear advantage for internalization. Licensing another firm to perform the entire production process in another location (as a substitute for horizontal FDI) often involves a substantial risk of losing some proprietary technology. On the other hand, there are no clear reasons why an independent firm should be able to replicate that production process at a lower cost than the parent firm. This gives internalization a strong advantage, so horizontal FDI is widely favored over the alternative of technology licensing to replicate the production process.

The trade-off between outsourcing and vertical FDI is much less clear-cut. There are many reasons why an independent firm could produce some parts of the production process at lower cost than the parent firm (in the same location). First and foremost, an independent firm can specialize in exactly that narrow part of the production process. As a result, it can also benefit from economies of scale if it performs those processes for many different parent firms.[21] Other reasons stress the advantages of local ownership in the alignment and monitoring of managerial incentives at the production facility.

But internalization also provides its own benefits when it comes to vertical integration between a firm and its supplier of a critical input to production: This avoids (or at least lessens) the potential for a costly renegotiation conflict after an initial agreement has been reached. Such conflicts can arise regarding many specific attributes of the input that cannot be specified in (or enforced by) a legal contract written at the time of the initial agreement. This can lead to a holdup of production by either party. For example, the buying firm can

---

[21]Companies that provide outsourced goods and services have expanded their list of clients to such an extent that they have now become large multinationals themselves. They specialize in providing a narrow set of services (or parts of the production process), but replicate this many times over for client companies across the globe.

claim that the quality of the part is not exactly as specified and demand a lower price. The supplying firm can claim that some changes demanded by the buyer led to increased costs and demand a higher price at delivery time.

Much progress has been made in recent research formalizing those trade-offs. This research explains how this important internalization choice is made, by describing when a firm chooses to integrate with its suppliers via vertical FDI and when it chooses an independent contractual relationship with those suppliers abroad. Developing those theories is beyond the scope of this textbook; ultimately, many of those theories boil down to different trade-offs between production cost savings and the fixed cost of moving parts of the production process abroad.

Describing which types of firms pick one offshoring option versus the other is sensitive to the details of the modeling assumptions. Nonetheless, one robust prediction emerges from those models when one compares either offshoring option to that of no offshoring (not breaking up the production chain and moving parts of it abroad). Relative to no offshoring, both vertical FDI and foreign outsourcing involve lower production costs combined with a higher fixed cost. As we saw, this implies a scale cutoff for a firm to choose either offshoring option. Thus, only the larger firms will choose either offshoring option and import some of their intermediate inputs.

This sorting scheme for firms to import intermediate goods is similar to the one we described for the firm's export choice: Only a subset of relatively more productive (lower-cost) firms will choose to offshore (import intermediate goods) and export (reach foreign customers)—because those are the firms that operate at sufficiently large scale to favor the trade-off involving higher fixed costs and lower per-unit costs (production- or trade-related).

Empirically, are the firms that offshore and import intermediate goods the same set of firms that also export? The answer is a resounding yes. For the United States in 2000, 92 percent of firms (weighed by employment) that imported intermediate goods also exported. Those importers thus also shared the same characteristics as U.S. exporters: They were substantially larger and more productive than the U.S. firms that did not engage in international trade.

Consequences of Multinationals and Foreign Outsourcing

Earlier in this chapter, we mentioned that internal economies of scale, product differentiation, and performance differences across firms combined to deliver some new channels for the gains from trade: increased product variety, and higher industry performance as firms move down their average cost curve and production is concentrated in the larger, more productive firms. What are the consequences for welfare of the expansion in multinational production and outsourcing?

We just saw how multinationals and firms that outsource take advantage of cost differentials that favor moving production (or parts thereof) to particular locations. In essence, this is very similar to the relocation of production that occurred *across* sectors when opening to trade. As we saw in Chapters 3 through 6, the location of production then shifts to take advantage of cost differences generated by comparative advantage.

We can therefore predict similar welfare consequences for the case of multinationals and outsourcing: Relocating production to take advantage of cost differences leads to overall gains from trade, but it is also likely to induce income distribution effects that leave some people worse off. We discussed one potential long-run consequence of outsourcing for income inequality in developed countries in Chapter 5.

Yet some of the most visible effects of multinationals and outsourcing occur in the short run, as some firms expand employment while others reduce employment in response

to increased globalization. We mentioned in Chapter 4 that those employment changes due to overseas plant relocations (along with plant closures due to import competition) account for only a small fraction (2.5 percent) of all involuntary worker displacements in the United States. Nevertheless, when such plant relocations do occur, they inevitably generate some substantial costs for those affected workers. As we argued in Chapter 4, the best policy response to this serious concern is still to provide an adequate safety net to unemployed workers without discriminating based on the economic force that induced their involuntary unemployment. Policies that impede firms' abilities to relocate production and take advantage of these cost differences may prevent these short-run costs for some, but they also forestall the accumulation of long-run economy-wide gains.

SUMMARY

1. Trade need not be the result of comparative advantage. Instead, it can result from increasing returns or economies of scale, that is, from a tendency of unit costs to be lower with larger output. Economies of scale give countries an incentive to specialize and trade even in the absence of differences between countries in their resources or technology. Economies of scale can be internal (depending on the size of the firm) or external (depending on the size of the industry).

2. Economies of scale internal to firms lead to a breakdown of perfect competition; models of imperfect competition must be used instead to analyze the consequences of increasing returns at the level of the firm. An important model of this kind is the monopolistic competition model, which is widely used to analyze models of firms and trade.

3. In monopolistic competition, an industry contains a number of firms producing differentiated products. These firms act as individual monopolists, but additional firms enter a profitable industry until monopoly profits are competed away. Equilibrium is affected by the size of the market: A large market will support a larger number of firms, each producing at a larger scale and thus a lower average cost, than a small market.

4. International trade allows for the creation of an integrated market that is larger than any one country's market. As a result, it is possible to simultaneously offer consumers a greater variety of products and lower prices. The type of trade generated by this model is intra-industry trade.

5. When firms differ in terms of their performance, economic integration generates winners and losers. The more productive (lower-cost) firms thrive and expand, while the less productive (higher-cost) firms contract. The least-productive firms are forced to exit.

6. In the presence of trade costs, markets are no longer perfectly integrated through trade. Firms can set different prices across markets. These prices reflect trade costs as well as the level of competition perceived by the firm. When there are trade costs, only a subset of more productive firms choose to export; the remaining firms serve only their domestic market.

7. Dumping occurs when a firm sets a lower price (net of trade costs) on exports than it charges domestically. A consequence of trade costs is that firms will feel competition more intensely on export markets because the firms have smaller market shares in those export markets. This leads firms to reduce markups for their export sales relative to their domestic sales; this behavior is characterized as dumping. Dumping is viewed as an unfair trade practice, but it arises naturally in a model of monopolistic competition and trade costs where firms from both countries behave in the same way. Policies against dumping are often used to discriminate against foreign firms in a market and erect barriers to trade.

8. Some multinationals replicate their production processes in foreign facilities located near large customer bases. This is categorized as horizontal foreign direct investment

(FDI). An alternative is to export to a market instead of operating a foreign affiliate in that market. The trade-off between exports and FDI involves a lower per-unit cost for FDI (no trade cost) but an additional fixed cost associated with the foreign facility. Only firms that operate at a big enough scale will choose the FDI option over exports.

9. Some multinationals break up their production chain and perform some parts of that chain in their foreign facilities. This is categorized as vertical foreign direct investment (FDI). One alternative is to outsource those parts of the production chain to an independent foreign firm. Both of those modes of operation are categorized as offshoring. Relative to the option of no offshoring, offshoring involves lower production costs but an additional fixed cost. Only firms that operate at a big enough scale will choose to offshore.

10. Multinational firms and firms that outsource parts of production to foreign countries take advantage of cost differences across production locations. This is similar to models of comparative advantage where production at the level of the industry is determined by differences in relative costs across countries. The welfare consequences are similar as well: There are aggregate gains from increased multinational production and outsourcing, but also changes in the income distribution that leaves some people worse off.

KEY TERMS

antidumping duty, p. 178
average cost, p. 158
dumping, p. 178
foreign direct investment
    (FDI), p. 180
foreign outsourcing, p. 185
horizontal FDI, p. 183
imperfect competition, p. 156

internal economies
    of scale, p. 155
internalization motive,
    p. 185
intra-industry trade, p. 169
location motive, p. 185
marginal cost, p. 158
marginal revenue, p. 157

markup over marginal
    cost, p. 163
monopolistic competition,
    p. 159
offshoring, p. 185
oligopoly, p. 159
pure monopoly, p. 157
vertical FDI, p. 183

PROBLEMS

1. In perfect competition, firms set price equal to marginal cost. Why can't firms do this when there are internal economies of scale?

2. Suppose the two countries we considered in the numerical example on pages 166–169 were to integrate their automobile market with a third country, which has an annual market for 3.75 million automobiles. Find the number of firms, the output per firm, and the price per automobile in the new integrated market after trade.

3. Suppose that fixed costs for a firm in the automobile industry (start-up costs of factories, capital equipment, and so on) are $5 billion and that variable costs are equal to $17,000 per finished automobile. Because more firms increase competition in the market, the market price falls as more firms enter an automobile market, or specifically, $P = 17,000 + (150/n)$, where n represents the number of firms in a market. Assume that the initial size of the U.S. and the European automobile markets are 300 million and 533 million people, respectively.

   a. Calculate the equilibrium number of firms in the U.S. and European automobile markets *without* trade.

   b. What is the equilibrium price of automobiles in the United States and Europe if the automobile industry is closed to foreign trade?

   c. Now suppose that the United States decides on free trade in automobiles with Europe. The trade agreement with the Europeans adds 533 million consumers to the automobile market, in addition to the 300 million in the United States. How

many automobile firms will there be in the United States and Europe combined? What will be the new equilibrium price of automobiles?

   d. Why are prices in the United States different in (c) and (b)? Are consumers better off with free trade? In what ways?

4. Go back to the model with firm performance differences in a single integrated market (pages 172–175). Now assume that a new technology becomes available. Any firm can adopt the new technology, but its use requires an additional fixed-cost investment. The benefit of the new technology is that it reduces a firm's marginal cost of production by a given amount.

   a. Could it be profit maximizing for some firms to adopt the new technology but not profit maximizing for other firms to adopt that same technology? Which firms would choose to adopt the new technology? How would they be different from the firms that choose not to adopt it?

   b. Now assume that there are also trade costs. In the new equilibrium with both trade costs and technology adoption, firms decide whether to export and also whether to adopt the new technology. Would exporting firms be more or less likely to adopt the new technology relative to nonexporters? Why?

5. In the chapter, we described a situation where dumping occurs between two symmetric countries. Briefly describe how things would change if the two countries had different sizes.

   a. How would the number of firms competing in a particular market affect the likelihood that an exporter to that market would be accused of dumping? (Assume that the likelihood of a dumping accusation is related to the firm's price difference between its domestic price and its export price: the higher the price difference, the more likely the dumping accusation.)

   b. Would a firm from a small country be more or less likely to be accused of dumping when it exports to a large country (relative to a firm from the large country exporting to the small country)?

6. Which of the following are direct foreign investments?

   a. A Saudi businessman buys $10 million of IBM stock.

   b. The same businessman buys a New York apartment building.

   c. A French company merges with an American company; stockholders in the U.S. company exchange their stock for shares in the French firm.

   d. An Italian firm builds a plant in Russia and manages the plant as a contractor to the Russian government.

7. For each of the following, specify whether the foreign direct investment is horizontal or vertical; in addition, describe whether that investment represents an FDI inflow or outflow from the countries that are mentioned.

   a. McDonald's (a U.S. multinational) opens up and operates new restaurants in Europe.

   b. Total (a French oil multinational) buys ownership and exploration rights to oil fields in Cameroon.

   c. Volkswagen (a German multinational auto producer) opens some new dealerships in the United States. (Note that, at this time, Volkswagen does not produce any cars in the United States.)

   d. Nestlé (a Swiss multinational producer of foods and drinks) builds a new production factory in Bulgaria to produce Kit Kat chocolate bars. (Kit Kat bars are produced by Nestlé in 17 countries around the world.)

8. If there are internal economies of scale, why would it ever make sense for a firm to produce the same good in more than one production facility?

9. Most firms in the apparel and footwear industries choose to outsource production to countries where labor is abundant (primarily, Southeast Asia and the Caribbean)—but those firms do not integrate with their suppliers there. On the other hand, firms in many

capital-intensive industries choose to integrate with their suppliers. What could be some differences between the labor-intensive apparel and footwear industries on the one hand and capital-intensive industries on the other hand that would explain these choices?

10. Consider the example of industries in the previous problem. What would those choices imply for the extent of *intra-firm* trade across industries? That is, in what industries would a greater proportion of trade occur within firms?

FURTHER READINGS

A. B. Bernard, J. B. Jensen, S. J. Redding, and P. K. Schott. "Firms in International Trade." *Journal of Economic Perspectives* 21 (Summer 2007), pp. 105–130. A nontechnical description of empirical patterns of trade at the firm level that focuses on U.S. firms.

A. B. Bernard, J. B. Jensen, and P. K. Schott. "Importers, Exporters, and Multinationals: A Portrait of Firms in the US that Trade Goods," in T. Dunne, J. B. Jensen, and M. J. Roberts, eds. *Producer Dynamics: New Evidence from Micro Data*. Chicago: University of Chicago Press, 2009. A nontechnical description of empirical patterns of trade at the firm level that focuses on U.S. firms and multinationals operating in the United States.

Robert Feenstra. "Integration of Trade and Disintegration of Production in the Global Economy." *Journal of Economic Perspectives* 12 (Fall 1998), pp. 32–50. A description of how the supply chain has been broken up into many processes that are then performed in different locations.

Gordon Hanson, Raymond Mataloni, and Matthew Slaughter. "Vertical Production Networks in Multinational Firms." *Review of Economics and Statistics* 87 (March 2005), pp. 664–678. An empirical description of vertical FDI patterns based on multinationals operating in the United States.

Keith Head. *Elements of Multinational Strategy*. New York: Springer, 2007. A recent textbook focused on multinationals.

Elhanan Helpman. "Trade, FDI, and the Organization of Firms." *Journal of Economic Literature* 44 (September 2006), pp. 589–630. A technical survey of recent research on models that incorporate firm performance differences, and on multinationals and outsourcing.

Elhanan Helpman. *Understanding Global Trade*. Cambridge, MA: Harvard University Press, forthcoming. A nontechnical book that covers both comparative advantage theories of trade and more recent theories of trade based on the firm.

Elhanan Helpman and Paul Krugman. *Market Structure and Foreign Trade*. Cambridge: MIT Press, 1985. A technical presentation of monopolistic competition and other models of trade with economies of scale.

Henryk Kierzkowski, ed. *Monopolistic Competition in International Trade*. Oxford: Clarendon Press, 1984. A collection of papers representing many of the leading researchers in imperfect competition and international trade.

James Markusen. "The Boundaries of Multinational Enterprises and the Theory of International Trade." *Journal of Economic Perspectives* 9 (Spring 1995), pp. 169–189. A nontechnical survey of models of trade and multinationals.

Thierry Mayer and Gianmarco I. P. Ottaviano. "The Happy Few: The Internationalisation of European Firms: New Facts Based on Firm-Level Evidence." *Intereconomics* 43 (May/June 2008), pp. 135–148.

MYECONLAB CAN HELP YOU GET A BETTER GRADE

If your exam were tomorrow, would you be ready? For each chapter, MyEconLab Practice Tests and Study Plans pinpoint which sections you have mastered and which ones you need to study. That way, you are more efficient with your study time, and you are better prepared for your exams.

APPENDIX TO CHAPTER 8

Determining Marginal Revenue

In our exposition of monopoly and monopolistic competition, we found it useful to have an algebraic statement of the marginal revenue faced by a firm given the demand curve it faced. Specifically, we asserted that if a firm faces the demand curve

$$Q = A - B \times P, \tag{8A-1}$$

its marginal revenue is

$$MR = P - (1/B) \times Q. \tag{8A-2}$$

In this appendix we demonstrate why this is true.

Notice first that the demand curve can be rearranged to state the price as a function of the firm's sales rather than the other way around. By rearranging (8A-1) we get

$$P = (A/B) - (1/B) \times Q. \tag{8A-3}$$

The revenue of a firm is simply the price it receives per unit multiplied by the number of units it sells. Letting R denote the firm's revenue, we have

$$R = P \times Q = [(A/B) - (1/B) \times Q] \times Q. \tag{8A-4}$$

Let us next ask how the revenue of a firm changes if it changes its sales. Suppose that the firm decides to increase its sales by a small amount, dX, so that the new level of sales is $Q = Q + dQ$. Then the firm's revenue after the increase in sales, R', will be

$$R' = P' \times Q' = [(A/B) - (1/B) \times (Q + dQ)] \times (Q + dQ) \tag{8A-5}$$
$$= [(A/B) - (1/B) \times Q] \times Q + [(A/B) - (1/B) \times Q] \times dQ$$
$$- (1/B) \times Q \times dQ - (1/B) \times (dQ)^2.$$

Equation (8A-5) can be simplified by substituting in from (8A-1) and (8A-4) to get

$$R' = R + P \times dQ - (1/B) \times Q \times dQ - (1/B) \times (dQ)^2. \tag{8A-6}$$

When the change in sales dQ is small, however, its square $(dQ)^2$ is very small (e.g., the square of 1 is 1, but the square of 1/10 is 1/100). So for a small change in Q, the last term in (8A-6) can be ignored. This gives us the result that the *change* in revenue from a small change in sales is

$$R' - R = [(P - (1/B) \times Q)] \times dQ. \tag{8A-7}$$

So the increase in revenue *per unit of additional sales*—which is the definition of marginal revenue—is

$$MR = (R' - R)/dQ = P - (1/B) \times Q,$$

which is just what we asserted in equation (8A-2).