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 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 $\overline{P}$ as given, we can rewrite the demand curve (8-5) in the form

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

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 $(S/n) + S \times b \times \overline{P}$ 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

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

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:

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

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:

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

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.

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.

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.

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.

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 $\overline{P} + [1/(b \times n)]$. 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

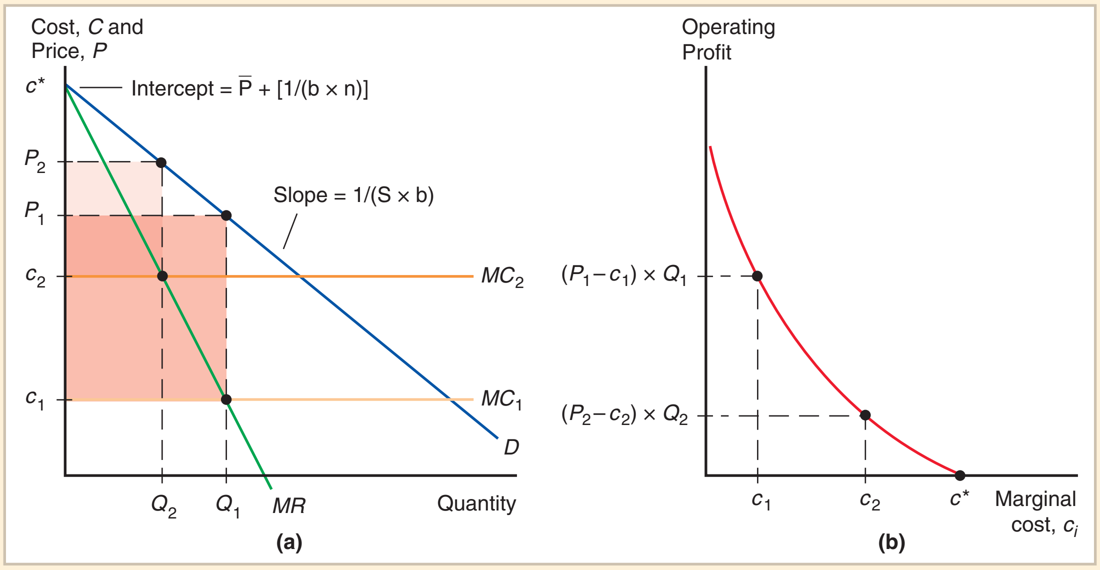

Figure 8-6

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 $\overline{P} + [1/(b \times n)]$. 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 $\overline{P} + [1/(b \times n)]$, 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 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.

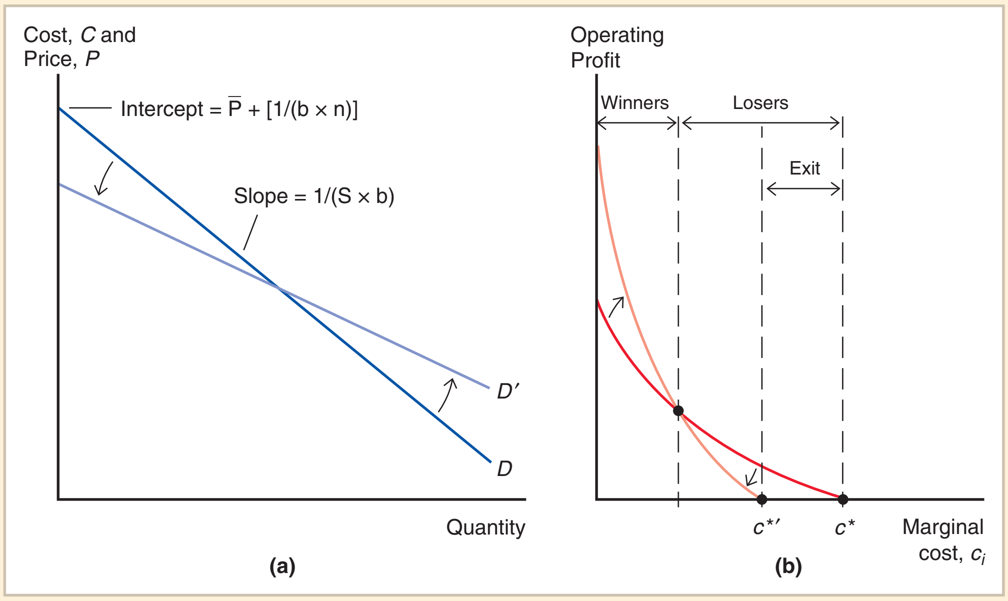

Figure 8-7

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, $c^{*\prime}$, 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.

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.

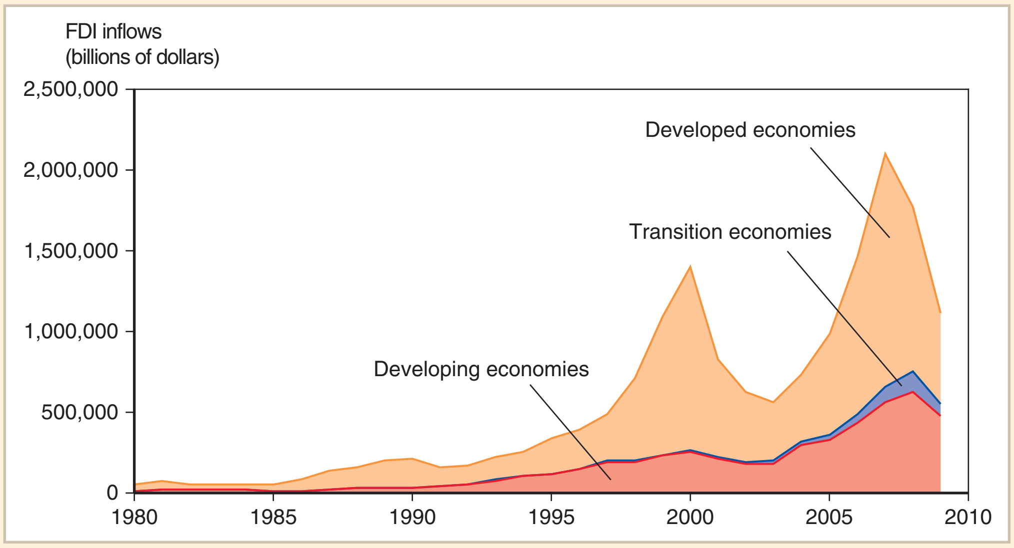

Figure 8-9

from spreadsheet 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 developed 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

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

(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

myeconlab

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.