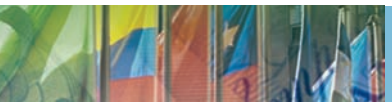


CHAPTER 6

Economies of Scale, Imperfect Competition, and International Trade

In Chapter 3 we pointed out that there are two reasons why countries specialize and trade. First, countries differ either in their resources or in technology and specialize in the things they do relatively well; second, economies of scale (or increasing returns) make it advantageous for each country to specialize in the production of only a limited range of goods and services. The past three chapters considered models in which all trade is based on comparative advantage; that is, differences between countries are the only reason for trade. This chapter introduces the role of economies of scale.

The analysis of trade based on economies of scale presents certain problems that we have so far avoided. Up to now we have assumed that markets are perfectly competitive, so that all monopoly profits are always competed away. When there are increasing returns, however, large firms usually have an advantage over small, so that markets tend to be dominated by one firm (monopoly) or, more often, by a few firms (oligopoly). When increasing returns enter the trade picture, then, markets usually become imperfectly competitive.

This chapter begins with an overview of the concept of economies of scale and the economics of imperfect competition. We then turn to two models of international trade in which economies of scale and imperfect competition play a crucial role: the monopolistic competition model and the dumping model. The rest of the chapter addresses the role of a different kind of increasing returns, external economies, in determining trade patterns.

Learning Goals

After reading this chapter, you will be able to:

- Recognize why international trade often occurs from increasing returns to scale and imperfect competition.
- Understand the source of *intra*industry trade and how it differs from *inter*industry trade.
- Detail the "dumping" arguments used by domestic industries as a basis for protectionism, and explain the relationship between dumping and price discrimination.
- Discuss the role of external economies and knowledge spillovers in shaping comparative advantage and international trade patterns.

Economies of Scale and International Trade: An Overview

The models of comparative advantage already presented were based on the assumption of constant returns to scale. That is, we assumed that if inputs to an industry were doubled, industry output would double as well. In practice, however, many industries are characterized by economies of scale (also referred to as increasing returns), so that production is more efficient the larger the scale at which it takes place. Where there are economies of scale, doubling the inputs to an industry will more than double the industry's production.

A simple example can help convey the significance of economies of scale for international trade. Table 6-1 shows the relationship between input and output of a hypothetical industry. Widgets are produced using only one input, labor; the table shows how the amount of labor required depends on the number of widgets produced. To produce 10 widgets, for example, requires 15 hours of labor, while to produce 25 widgets requires 30 hours. The presence of economies of scale may be seen from the fact that doubling the input of labor from 15 to 30 more than doubles the industry's output—in fact, output increases by a factor of 2.5. Equivalently, the existence of economies of scale may be seen by looking at the average amount of labor used to produce each unit of output: If output is only 5 widgets, the average labor input per widget is 2 hours, while if output is 25 units, the average labor input falls to 1.2 hours.

We can use this example to see why economies of scale provide an incentive for international trade. Imagine a world consisting of two countries, America and Britain, both of which have the same technology for producing widgets, and suppose that initially each country produces 10 widgets. According to the table, this requires 15 hours of labor in each country, so in the world as a whole 30 hours of labor produce 20 widgets. But now

TABLE 6-1 Relationship of Input to Output for a Hypothetical Industry

Output	Total Labor Input	Average Labor Input
5	10	2
10	15	1.5
15	20	1.333333
20	25	1.25
25	30	1.2
30	35	1.166667

suppose that we concentrate world production of widgets in one country, say America, and let America employ 30 hours of labor in the widget industry. In a single country these 30 hours of labor can produce 25 widgets. So by concentrating production of widgets in America, the world economy can use the same amount of labor to produce 25 percent more widgets.

But where does America find the extra labor to produce widgets, and what happens to the labor that was employed in the British widget industry? To get the labor to expand its production of some goods, America must decrease or abandon the production of others; these goods will then be produced in Britain instead, using the labor formerly employed in the industries whose production has expanded in America. Imagine that there are many goods subject to economies of scale in production, and give them numbers: 1, 2, 3, ... To take advantage of economics of scale, each of the countries must concentrate on producing only a limited number of goods. Thus, for example, America might produce goods 1, 3, 5, and so on while Britain produces 2, 4, 6, and so on. If each country produces only some of the goods, then each good can be produced at a larger scale than would be the case if each country tried to produce everything, and the world economy can therefore produce more of each good.

How does international trade enter the story? Consumers in each country will still want to consume a variety of goods. Suppose that industry 1 ends up in America and industry 2 in Britain; then American consumers of good 2 will have to buy goods imported from Britain, while British consumers of good 1 will have to import it from America. International trade plays a crucial role: It makes it possible for each country to produce a restricted range of goods and to take advantage of economies of scale without sacrificing variety in consumption. Indeed, as we will see below, international trade typically leads to an increase in the variety of goods available.

Our example, then, suggests how mutually beneficial trade can arise as a result of economics of scale. Each country specializes in producing a limited range of products, which enables it to produce these goods more efficiently than if it tried to produce everything for itself; these specialized economies then trade with each other to be able to consume the full range of goods.

Unfortunately, to go from this suggestive story to an explicit model of trade based on economies of scale is not that simple. The reason is that economies of scale typically lead to a market structure other than that of perfect competition, and it is necessary to be careful about analyzing this market structure.

Economies of Scale and Market Structure

In the example in Table 6-1, we represented economies of scale by assuming that the labor input per unit of production is smaller the more units produced. We did not say how this production increase was achieved—whether existing firms simply produced more, or whether there was instead an increase in the number of firms. To analyze the effects of economies of scale on market structure, however, one must be clear about what kind of production increase is necessary to reduce average cost. **External economies of scale** occur when the cost per unit depends on the size of the industry but not necessarily on the size of any one firm. **Internal economies of scale** occur when the cost per unit depends on the size of an individual firm but not necessarily on that of the industry.

The distinction between external and internal economies can be illustrated with a hypothetical example. Imagine an industry that initially consists of ten firms, each producing 100 widgets, for a total industry production of 1,000 widgets. Now consider two cases. First, suppose the industry were to double in size, so that it now consists of 20 firms, each

one still producing 100 widgets. It is possible that the costs of each firm will fall as a result of the increased size of the industry; for example, a bigger industry may allow more efficient provision of specialized services or machinery. If this is the case, the industry exhibits external economies of scale. That is, the efficiency of firms is increased by having a larger industry, even though each firm is the same size as before.

Second, suppose the industry's output were held constant at 1,000 widgets, but that the number of firms is cut in half so that each of the remaining five firms produces 200 widgets. If the costs of production fall in this case, then there are internal economies of scale: A firm is more efficient if its output is larger.

External and internal economies of scale have different implications for the structure of industries. An industry where economies of scale are purely external (that is, where there are no advantages to large firms) will typically consist of many small firms and be perfectly competitive. Internal economics of scale, by contrast, give large firms a cost advantage over small and lead to an imperfectly competitive market structure.

Both external and internal economies of scale are important causes of international trade. Because they have different implications for market structure, however, it is difficult to discuss both types of scale economy-based trade in the same model. We will therefore deal with them one at a time.

We begin with a model based on internal economics of scale. As we have just argued, however, internal economies of scale lead to a breakdown of perfect competition. This outcome forces us to take time out to review the economics of imperfect competition before we can turn to the analysis of the role of internal economics of scale in international trade.

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, sellers of products believe that they can sell as much as they like at the current price and 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, matters are 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. Boeing therefore 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 the other way around, 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. Imperfect competition is characteristic both of industries in which there are only a few major producers and of industries in which each producer's product is seen by consumers as strongly differentiated from those of rival firms. 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 can then be used to examine more complex market structures.

Monopoly: A Brief Review

Figure 6-1 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 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 and will therefore not have to lower the price on units he would have sold otherwise by very much, 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 marginal revenue much less than price.

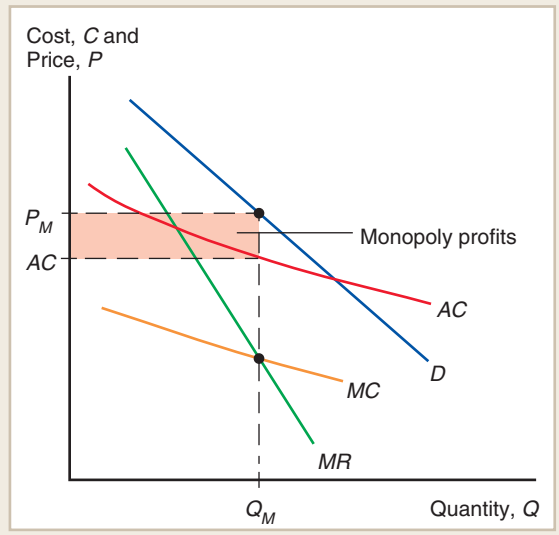
We can be more specific about the relationship between price and marginal revenue if we assume that the demand curve the firm faces is a straight line. When this is so, the dependence of the monopolist's total sales on the price it charges can be represented by an equation of the form

$$Q = A - B \times P, \tag{6-1}$$

Figure 6-1

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



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

$$\text{Marginal revenue} = MR = P - Q/B, \tag{6-2}$$

implying

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

Equation (6-2) 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. The greater is B, that is, the more sales fall for any given increase in price and the closer marginal revenue is to the price of the good. Equation (6-2) is crucial for our analysis of the monopolistic competition model of trade (pp. 125–143).

Average and Marginal Costs Returning to Figure 6-1, 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 economics of scale, so that the larger the firm's output is the lower are its costs per unit. MC represents the firm's **marginal cost** (the amount it costs the firm to produce one extra unit). We know from basic economics that when average costs are a decreasing function of output, marginal cost is always less than average cost. Thus MC lies below AC.

Equation (6-2) related price and marginal revenue. There is a corresponding formula relating average and marginal cost. Suppose the costs of a firm, C, take the form

$$C = F + c \times Q, \tag{6-3}$$

where F is a fixed cost that is independent of the firm's output, c is the firm's marginal cost, and Q is once again the firm's output. (This is called a linear cost function.) *The fixed cost in a linear cost function gives rise to economies of scale, because the larger the firm's output, the less is the fixed cost per unit.* Specifically, the firm's average cost (total cost divided by output) is

$$\text{Average cost} = AC = C/Q = F/Q + c. \tag{6-4}$$

This average cost declines as Q increases because the fixed cost is spread over a larger output.

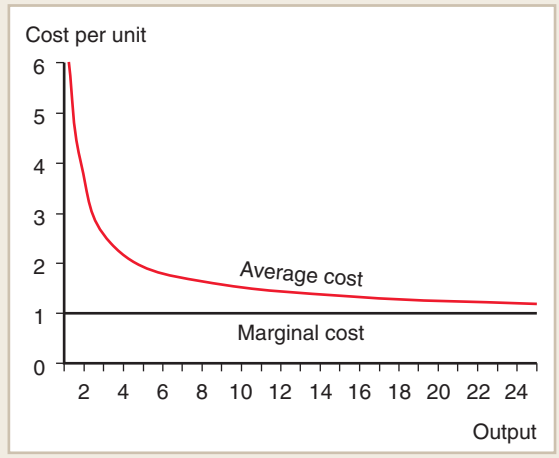
If, for example, $F = 5$ and $c = 1$ the average cost of producing 10 units is $5/10 + 1 = 1.5$ and the average cost of producing 25 units is $5/25 + 1 = 1.2$. These numbers may look familiar, because they were used to construct Table 6-1. The relationship between output, average costs, and marginal costs given in Table 6-1 is shown graphically in Figure 6-2. 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 6-1 we can see that the price at which the profit-maximizing output Q_M is demanded is P_M, which

Figure 6-2

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.



is greater than average cost. When $P > AC$, the monopolist is earning some monopoly profits.[1]

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:** several firms, each of them large enough to affect prices, but none with 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 others' strategies. We will briefly discuss the general problems of modeling oligopoly below. However, there is a special case of oligopoly, known as monopolistic competition, which is relatively easy to analyze. Since 1980 monopolistic competition models have been widely applied to 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 they want to buy this firm's particular product, the firm's customers will not rush to buy other firms' products because of a slight price difference. Product differentiation assures 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, it behaves as if it were a monopolist—hence the model's name.

[1] 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.

Are there any monopolistically competitive industries in the real world? Some industries may be reasonable approximations. For example, the automobile industry in Europe, where a number of major producers (Ford, General Motors, Volkswagen, Renault, Peugeot, Fiat, Volvo—and more recently Nissan) offer substantially different yet nonetheless competing automobiles, may be fairly well described by monopolistically competitive assumptions. The main appeal of the monopolistic competition model is not, however, 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 number of firms. These firms produce differentiated products, that is, goods that are not exactly the same but that are 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 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 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[2]

$$Q = S \times [1/n - b \times (P - \overline{P})], \tag{6-5}$$

where Q is the firm's sales, S is the total sales of the industry, n the number of firms in the industry, b a constant term representing the responsiveness of a firm's sales to its price, P the price charged by the firm itself, and $\overline{P}$ the average price charged by its competitors. Equation (6-5) 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, a firm charging less a larger share.[3]

It is helpful to assume that total industry sales S are 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 others' expense. This is an unrealistic assumption, but it simplifies the analysis and helps 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 (6-3) and (6-4).

Market Equilibrium To model the behavior of this monopolistically competitive industry, we will assume that all firms in this industry are *symmetric,* that is, the demand function and cost function are identical for all firms (even though they are producing and selling somewhat differentiated products). When the individual firms are symmetric, the state of

[2] Equation (6-5) 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.

[3] Equation (6-5) may be rewritten as $Q = S/n - S \times b \times (P - \overline{P})$. If $P = \overline{P}$, this reduces to $Q = S/n$. If $P > \overline{P}$, $Q < S/n$, while if $P < \overline{P}$, $Q > S/n$.

the industry can be described without enumerating the features of all firms in detail: 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 its 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 competition among firms, and as a result the lower the prices they charge. (3) Finally, we argue that when the price exceeds average cost additional firms will enter the industry, while when the price is less than average cost firms will exit. So in the long run 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 (6-5) 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 (6-4) 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:

$$AC = F/Q + c = n \times F/S + c. \tag{6-6}$$

Equation (6-6) 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 6-3.

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 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 (6-1), and then to use equation (6-2) to determine prices.

First recall that in the monopolistic competition model firms are assumed to take each others' 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 (6-5) in the form

$$Q = (S/n + S \times b \times \overline{P}) - S \times b \times P, \tag{6-7}$$

where b is the parameter in equation (6-5) that measured the sensitivity of each firm's market share to the price it charges. Now this is in the same form as (6-1), 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

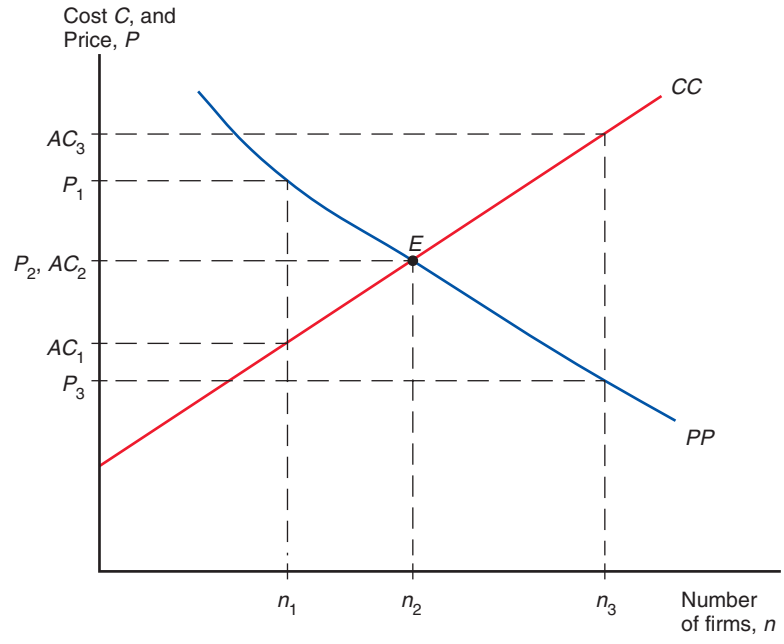


Figure 6-3

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 its average cost. This relationship is represented by *CC*. If price exceeds average cost (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*.

coefficient B. If we plug these values back into the formula for marginal revenue (6-2), we have a marginal revenue for a typical firm of

$$MR = P - Q/(S \times b). \tag{6-8}$$

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:

$$P = c + Q/(S \times b). \tag{6-9}$$

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 (6-9) gives us a relationship between the number of firms and the price each firm charges:

$$P = c + 1/(b \times n). \tag{6-10}$$

Equation (6-10) says algebraically that *the more firms there are in the industry, the lower the price each firm will charge.* Equation (6-10) is shown in Figure 6-3 as the downward-sloping curve *PP*.

3. *The equilibrium number of firms.* Let us now ask what Figure 6-3 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, 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 P_2 the equilibrium price.[4]

We have now 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 can use this model to derive some important conclusions about the role of economies of scale in international trade. But before we do, we should take a moment to note some limitations of the monopolistic competition model.

Limitations of the Monopolistic Competition Model

The monopolistic competition model captures certain key elements of markets where there are economies of scale and thus imperfect competition. However, few industries are well described by monopolistic competition. Instead, the most common market structure is one of small-group oligopoly, where only a few firms are actively engaged in competition. In this situation the key assumption of the monopolistic competition model, which is that each firm will behave as if it were a true monopolist, is likely to break down. Instead, firms

[4] 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, but not 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.

will be aware that their actions influence the actions of other firms and will take this interdependence into account.

Two kinds of behavior arise in the general oligopoly setting that are excluded by assumption from the monopolistic competition model. The first is *collusive* behavior. Each firm may keep its price higher than the apparent profit-maximizing level as part of an understanding that other firms will do the same; since each firm's profits are higher if its competitors charge high prices, such an understanding can raise the profits of all the firms (at the expense of consumers). Collusive price-setting behavior may be managed through explicit agreements (illegal in the United States) or through tacit coordination strategies, such as allowing one firm to act as a price leader for the industry.

Firms may also engage in *strategic* behavior; that is, they may do things that seem to lower profits but that affect the behavior of competitors in a desirable way. For example, a firm may build extra capacity not to use it but to deter potential rivals from entering its industry.

These possibilities for both collusive and strategic behavior make the analysis of oligopoly a complex matter. There is no one generally accepted model of oligopoly behavior, which makes modeling trade in monopolistic industries problematic.

The monopolistic competition approach to trade is attractive because it avoids these complexities. Even though it may leave out some features of the real world, the monopolistic competition model is widely accepted as a way to provide at least a first cut at the role of economics 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 specialize in producing a narrower range of products than it would in the absence of trade; yet by buying goods that it does not make from other countries, 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.

To see this in the context of our model, look again at the *CC* curve in Figure 6-3, 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 (6-6):

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

Examining this equation, we see that an increase in total sales 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, sales 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 6-3, which relates the price charged by firms to the number of firms, does not shift. The definition of that curve was given in equation (6-10):

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

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

Figure 6-4 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

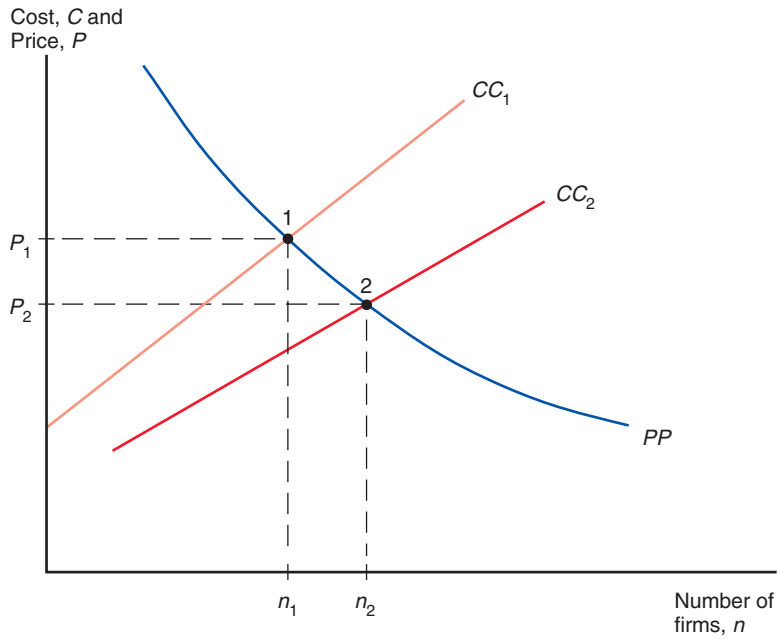


Figure 6-4

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 fall in the price of each.

firms n_1. An increase in the size of the market, measured by industry sales S, shifts 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 (6-5), 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

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

where Q is the number of automobiles sold per firm, S the total sales of the industry, n the number of firms, P the price that a firm charges, and $\overline{P}$ the average price of other firms. We also assume that the cost function for producing automobiles is described by equation (6-3), with a fixed cost $F = \$750{,}000{,}000$ and a marginal cost $c = \$5{,}000$ per automobile (again these values are 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 6-5a 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 (6-10) 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 (6-10), we find

$$\begin{aligned} P = \$10{,}000 = c + 1/(b \times n) &= \$5{,}000 + 1/[(1/30{,}000) \times 6] \\ &= \$5{,}000 + \$5{,}000, \end{aligned}$$

so the condition for profit maximization—that marginal revenue equal 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.

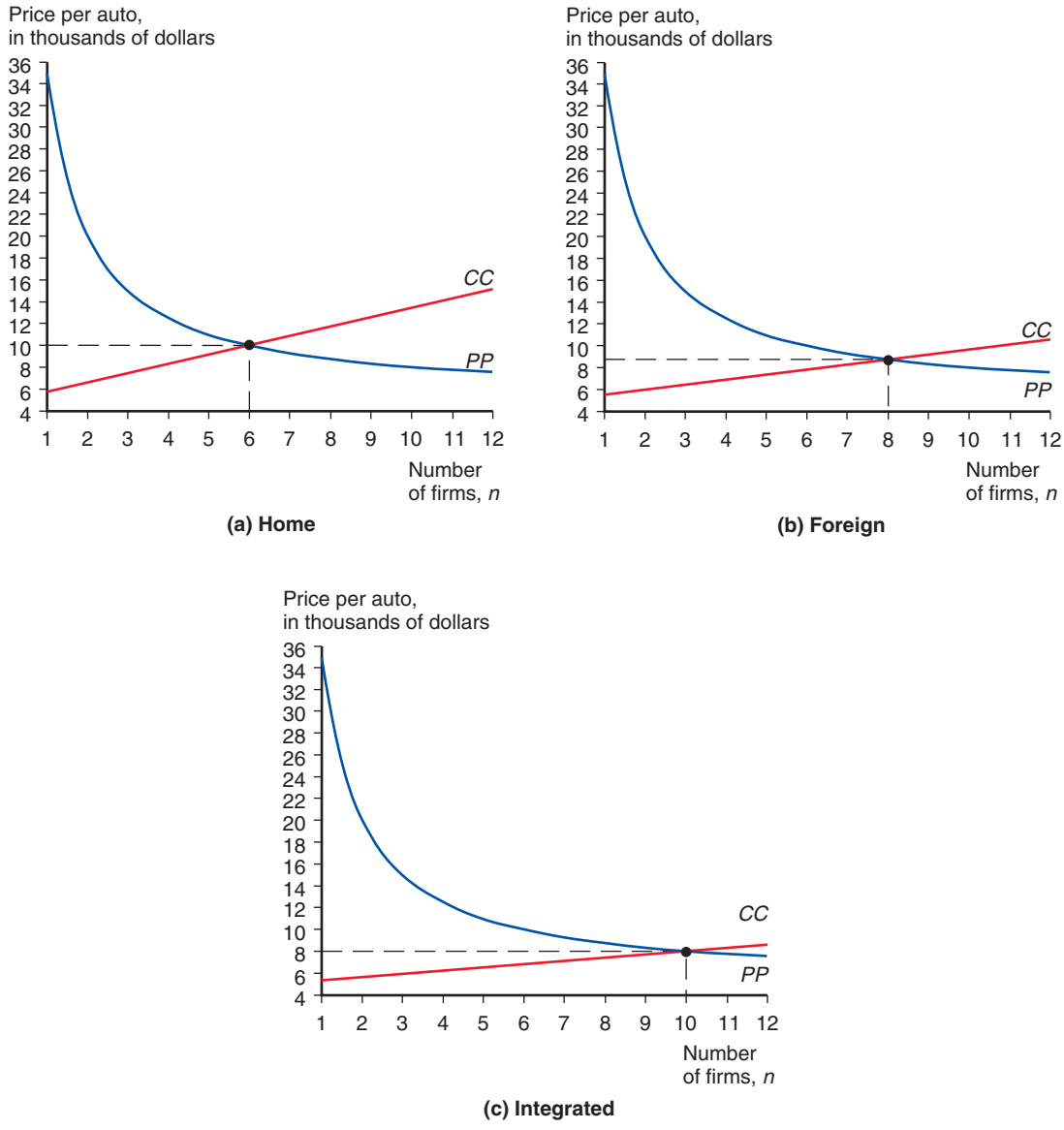


Figure 6-5

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 car. (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 6-5) 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 6-5) 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 6-2. 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 did on its own.

Clearly everyone is better off as a result of integration. In the larger market, consumers have a wider range of choice, 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.

Economies of Scale and Comparative Advantage

Our example of a monopolistically competitive industry says little about the pattern of trade that results from economies of scale. The model assumes that the cost of production is the same in both countries and that trade is costless. These assumptions mean that although we know that the integrated market will support ten firms, we cannot say where they will be located. For example, four firms might be in Home and six in Foreign—but it is equally possible, as far as this example goes, that all ten will be in Foreign (or in Home).

To say more than that the market will support ten firms, it is necessary to go behind the partial equilibrium framework that we have considered so far and think about how economies of scale interact with comparative advantage to determine the pattern of international trade.

Let us therefore now imagine a world economy consisting, as usual, of our two countries Home and Foreign. Each of these countries has two factors of production, capital and

TABLE 6-2 Hypothetical Example of Gains from Market Integration

	Home Market, Before Trade	Foreign Market, Before Trade	Integrated Market, After Trade
Total sales of autos	900,000	1,600,000	2,500,000
Number of firms	6	8	10
Sales per firm	150,000	200,000	250,000
Average cost	$10,000	$8,750	$8,000
Price	$10,000	$8,750	$8,000

labor. We assume that Home has a higher overall capital-labor ratio than Foreign, that is, that Home is the capital-abundant country. Let's also imagine that there are two industries, cloth and food, with cloth the more capital-intensive industry.

The difference between this model and the factor proportions model of Chapter 4 is that we now suppose that cloth is not a perfectly competitive industry producing a homogeneous product. Instead, it is a monopolistically competitive industry in which a number of firms all produce differentiated products. *Because of economies of scale, neither country is able to produce the full range of cloth products by itself; thus, although both countries may produce some cloth, they will be producing different things.* The monopolistically competitive nature of the cloth industry makes an important difference to the trade pattern, a difference that can best be seen by looking at what would happen if cloth were *not* a monopolistically competitive sector.

If cloth were *not* a differentiated product sector, we know from Chapter 4 what the trade pattern would look like. Because Home is capital-abundant and cloth capital-intensive, Home would have a larger relative supply of cloth and would therefore export cloth and import food. Schematically, we can represent this trade pattern with a diagram like Figure 6-6. The length of the arrows indicates the value of trade in each direction; the figure shows that Home would export cloth equal in value to the food it imports.

If we assume that cloth is a monopolistically competitive sector (each firm's products are differentiated from other firms'), Home will still be a *net* exporter of cloth and an importer of food. However, Foreign firms in the cloth sector will produce products different from those that Home firms produce. Because some Home consumers will prefer Foreign varieties, Home, although running a trade surplus in cloth, will import as well as export within the cloth industry. With cloth monopolistically competitive, then, the pattern of trade will look like Figure 6-7.

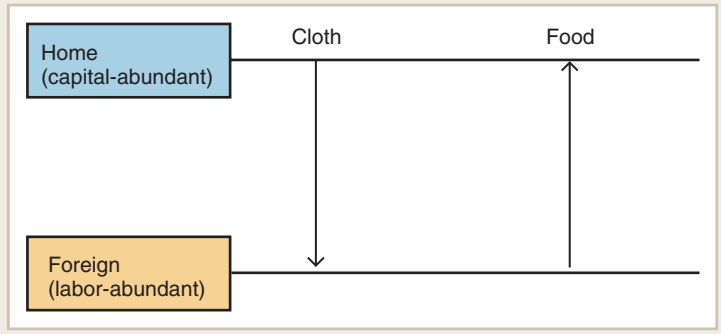


Figure 6-6
Trade in a World Without Increasing Returns
In a world without economies of scale, there would be a simple exchange of cloth for food.

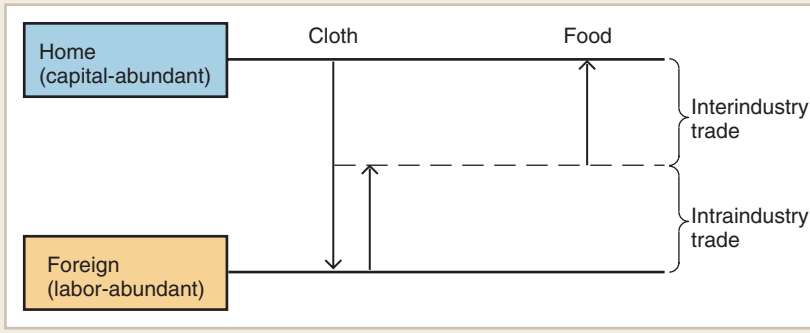


Figure 6-7

Trade with Increasing Returns and Monopolistic Competition

If cloth is a monopolistically competitive industry, Home and Foreign will produce differentiated products. As a result, even if Home is a net exporter of cloth goods, it will import as well as export cloth, giving rise to intraindustry trade.

We can think of world trade in a monopolistic competition model as consisting of two parts. There will be two-way trade *within* the cloth sector. This exchange of cloth for cloth is called **intraindustry trade.** The remainder of trade is an exchange of cloth for food called **interindustry trade.**

Notice these four points about this pattern of trade:

1. *Interindustry* (cloth for food) trade reflects comparative advantage. The pattern of interindustry trade is that Home, the capital-abundant country, is a net exporter of capital-intensive cloth and a net importer of labor-intensive food. So comparative advantage continues to be a major part of the trade story.

2. *Intraindustry* trade (cloth for cloth) does *not* reflect comparative advantage. Even if the countries had the same overall capital-labor ratio, their firms would continue to produce differentiated products and the demand of consumers for products made abroad would continue to generate intraindustry trade. It is economies of scale that keep each country from producing the full range of products for itself; thus economies of scale can be an independent source of international trade.

3. The pattern of intraindustry trade itself is unpredictable. We have not said anything about which country produces which goods within the cloth sector because there is nothing in the model to tell us. All we know is that the countries will produce different products. Since history and accident determine the details of the trade pattern, an unpredictable component of the trade pattern is an inevitable feature of a world where economies of scale are important. Notice, however, that the unpredictability is not total. While the precise pattern of intraindustry trade within the cloth sector is arbitrary, the pattern of interindustry trade between cloth and food is determined by underlying differences between countries.

4. The relative importance of intraindustry and interindustry trade depends on how similar countries are. If Home and Foreign are similar in their capital-labor ratios, then there will be little interindustry trade, and intraindustry trade, based ultimately on economies of scale, will be dominant. On the other hand, if the capital-labor ratios are very different, so that, for example, Foreign specializes completely in food production, there will be no intraindustry trade based on economies of scale. All trade will be based on comparative advantage.

The Significance of Intraindustry Trade

About one-fourth of world trade consists of intraindustry trade, that is, two-way exchanges of goods within standard industrial classifications. Intraindustry trade plays a particularly large role in the trade in manufactured goods among advanced industrial nations, which accounts for most of world trade. Over time, the industrial countries have become increasingly similar in their levels of technology and in the availability of capital and skilled labor. Since the major trading nations have become similar in technology and resources, there is often no clear comparative advantage within an industry, and much of international trade therefore takes the form of two-way exchanges within industries—probably driven in large part by economies of scale—rather than interindustry specialization driven by comparative advantage.

Table 6-3 shows measures of the importance of intraindustry trade for a number of U.S. manufacturing industries in 1993. The measure shown is intraindustry trade/total trade.[5] The measure ranges from 0.99 for inorganic chemicals—an industry in which U.S. exports and imports are nearly equal—to 0.00 for footwear, an industry in which the United States has large imports but virtually no exports. The measure would be zero for an industry in which the United States was only an exporter or only an importer, not both; it would be one in an industry for which U.S. exports exactly equaled U.S. imports.

Table 6-3 shows that in many industries a large part of trade is intraindustry (closer to one) rather than interindustry (closer to zero). The industries are ranked by the relative importance of intraindustry trade, those with higher intraindustry trade first. Industries with high levels of intraindustry trade tend to be sophisticated manufactured goods, such as chemicals, pharmaceuticals, and power-generating equipment. These goods are

TABLE 6-3 Indexes of Intraindustry Trade for U.S. Industries, 1993

Inorganic chemicals	0.99
Power-generating machinery	0.97
Electrical machinery	0.96
Organic chemicals	0.91
Medical and pharmaceutical	0.86
Office machinery	0.81
Telecommunications equipment	0.69
Road vehicles	0.65
Iron and steel	0.43
Clothing and apparel	0.27
Footwear	0.00

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

$$I = 1 - \frac{|\text{exports} - \text{imports}|}{\text{exports} + \text{imports}}$$

where the expression |exports–imports| means "absolute value of the trade balance": If exports are $100 million more than imports, the numerator of the fraction is 100, but if exports are $100 million *less* than imports, it is also 100. In comparative advantage models of international trade, we expect a country either to export a good or to import it, not both; in that case I would always equal zero. On the other hand, if a country's exports and imports within an industry are equal, we find $I = 1$.

exported principally by advanced nations and are probably subject to important economies of scale in production. At the other end of the scale, the industries with very little intraindustry trade are typically 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 clear-cut and is the primary determinant of U.S. trade with these countries.[6]

Why Intraindustry Trade Matters

Table 6-3 shows that a sizeable part of international trade is intraindustry trade rather than the interindustry trade we studied in Chapters 3 through 5. But does the importance of intraindustry trade change any of our conclusions?

First, intraindustry trade produces extra gains from international trade, over and above those from comparative advantage, because intraindustry trade allows countries to benefit from larger markets. As we have seen, by engaging in intraindustry trade a country can simultaneously reduce the number of products it produces *and* increase the variety of goods available to domestic consumers. By producing fewer varieties, a country can produce each at larger scale, with higher productivity and lower costs. At the same time, consumers benefit from the increased range of choice. In our numerical example of the gains from integrating a market, Home consumers found that intraindustry trade expanded their range of choice from six automobile models to ten even as it reduced the price of autos from $10,000 to $8,000. As the case study of the North American auto industry indicates (p. 134), the advantages of creating an integrated industry in two countries can be substantial in reality as well.

In our earlier analysis of the distribution of gains from trade (Chapter 4), we were pessimistic about the prospects that everyone will benefit from trade, even though international trade could potentially raise everyone's income. In the models discussed earlier, trade had all its effects through changes in relative prices, which in turn have very strong effects on the distribution of income.

Suppose, however, that intraindustry trade is the dominant source of gains from trade. This will happen (1) when countries are similar in their relative factor supplies, so that there is not much interindustry trade, and (2) when scale economies and product differentiation are important, so that the gains from larger scale and increased choice are large. In these circumstances the income distribution effects of trade will be small and there will be substantial extra gains from intraindustry trade. The result may well be that despite the effects of trade on income distribution, everyone gains from trade.

When will this be most likely to happen? Intraindustry trade tends to be prevalent between countries that are similar in their capital-labor ratios, skill levels, and so on. Thus, intraindustry trade will be dominant between countries at a similar level of economic development. Gains from this trade will be large when economies of scale are strong and products are highly differentiated. This is more characteristic of sophisticated manufactured goods than of raw materials or more traditional sectors (such as textiles or footwear).

---

[6] The growing trade between low-wage and high-wage nations sometimes produces trade that is classified as intraindustry even though it is really driven by comparative advantage. Suppose, for example, a U.S. company produces some sophisticated computer chips in California, ships them to Asia where they are assembled into a computer, and then ships that computer back home. Both the exported components and the imported computer are likely to be classified as being "computers and related devices," so that the transactions will be counted as intraindustry trade. Nonetheless, what is really going on is that the United States is exporting skill-intensive products (chips) and importing a labor-intensive service (computer assembly). Such "pseudo-intraindustry" trade is particularly common in trade between the United States and Mexico.

Trade without serious income distribution effects, then, is most likely to happen in manufactures trade between advanced industrial countries.

This conclusion is borne out by postwar experience, particularly in Western Europe. In 1957 the major countries of continental Europe established a free trade area in manufactured goods, 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. Trade within the EEC grew twice as fast as world trade as a whole during the 1960s. One might have expected this rapid growth in trade to produce substantial dislocations and political problems. The growth in trade, however, was almost entirely intraindustry rather than interindustry; drastic economic dislocation did not occur. Instead of, say, workers in France's electrical machinery industry being hurt while those in Germany's gained, workers in both sectors gained from the increased efficiency of the integrated European industry. The result was that the growth in trade within Europe presented far fewer social and political problems than anyone anticipated.

There is both a good and a bad side to this favorable view of intraindustry trade. The good side is that under some circumstances trade is relatively easy to live with and therefore relatively easy to support politically. The bad side is that trade between very different countries or where scale economies and product differentiation are not important remains politically problematic. In fact, the progressive liberalization of trade that characterized the 30-year period from 1950 to 1980 was primarily concentrated on trade in manufactures among the advanced nations, as we will see in Chapter 9. If progress on other kinds of trade is important, the past record does not give us much encouragement.

Case Study

Intraindustry 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, 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 departure from our model, since we have not yet examined the role of multinational firms—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 important, 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 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. This was achieved by importing from the United States products no longer made in Canada and exporting the products Canada continued to make. 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: intraindustry 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.

Dumping

The monopolistic competition model helps us understand how increasing returns promote international trade. As we noted earlier, however, this model assumes away many of the issues that can arise when firms are imperfectly competitive. Although it recognizes that imperfect competition is a necessary consequence of economies of scale, the monopolistic competition analysis does not focus on the possible consequences of imperfect competition itself for international trade.

In reality, imperfect competition has some important consequences for international trade. The most striking of these is that firms do not necessarily charge the same price for goods that are exported and those that are sold to domestic buyers.

The Economics of Dumping

In imperfectly competitive markets, firms sometimes charge one price for a good when that good is exported and a different price for the same good when it is sold domestically. In general, the practice of charging different customers different prices is called **price discrimination.** The most common form of price discrimination in international trade is **dumping,** a pricing practice in which a firm charges a lower price for exported goods than it does for the same goods sold domestically. Dumping is a controversial issue in trade policy, where it is widely regarded as an "unfair" practice and is subject to special rules and penalties. We will discuss the policy dispute surrounding dumping in Chapter 9. For now, we present some basic economic analysis of the dumping phenomenon.

Dumping can occur only if two conditions are met. First, the industry must be imperfectly competitive, so that firms set prices rather than taking market prices as given. Second, markets must be *segmented,* so that domestic residents cannot easily purchase goods intended for export. Given these conditions, a monopolistic firm may find that it is profitable to engage in dumping.

An example may help to show how dumping can be a profit-maximizing strategy. Imagine a firm that currently sells 1,000 units of a good at home and 100 units abroad. Currently selling the good at $20 per unit domestically, it gets only $15 per unit on export sales. One might imagine that the firm would conclude that additional domestic sales are much more profitable than additional exports.

Suppose, however, that to expand sales by one unit, in either market, would require reducing the price by $0.01. Reducing the domestic price by a penny, then, would

increase sales by one unit—directly adding $19.99 in revenue, but reducing the receipts on the 1,000 units that would have sold at the $20 price by $10. So the marginal revenue from the extra unit sold is only $9.99. On the other hand, reducing the price charged to foreign customers and thereby expanding exports by one unit would directly increase revenue by only $14.99. The indirect cost of reduced receipts on the 100 units that would have been sold at the original price, however, would be only $1, so that marginal revenue on export sales would be $13.99. It would therefore be more profitable in this case to expand exports rather than domestic sales, even though the price received on exports is lower.

This example could be reversed, with the incentive being to charge less on domestic than foreign sales. However, price discrimination in favor of exports is more common. Since international markets are imperfectly integrated due to both transportation costs and protectionist trade barriers, domestic firms usually have a larger share of home markets than they do of foreign markets. This in turn usually means that their foreign sales are more affected by their pricing than their domestic sales. A firm with a 20 percent market share need not cut its price as much to double its sales as a firm with an 80 percent share. So firms typically see themselves as having less monopoly power, and a greater incentive to keep their prices low, on exports than on domestic sales.

Figure 6-8 offers a diagrammatic example of dumping. It shows an industry in which there is a single monopolistic domestic firm. The firm sells in two markets: a domestic market, where it faces the demand curve D_{DOM}, and an export market. In the export market we take the assumption that sales are highly responsive to the price the firm charges to an extreme, assuming the firm can sell as much as it wants at the price P_{FOR}. The horizontal line P_{FOR} is thus the demand curve for sales in the foreign market. We assume the markets are segmented, so that the firm can charge a higher price for domestically sold goods than it does for exports. *MC* is the marginal cost curve for total output, which can be sold on either market.

To maximize profits, the firm must set marginal revenue equal to marginal cost in *each* market. Marginal revenue on domestic sales is defined by the curve MR_{DOM}, which lies below D_{DOM}. Export sales take place at a constant price P_{FOR}, so the marginal revenue for an additional unit exported is just P_{FOR}. To set marginal cost equal to marginal revenue in both markets it is necessary to produce the quantity $Q_{MONOPOLY}$, to sell Q_{DOM} on the domestic market, and to export $Q_{MONOPOLY} - Q_{DOM}$.[7] The cost of producing an additional unit in this case is equal to P_{FOR}, the marginal revenue from exports, which in turn is equal to the marginal revenue for domestic sales.

The quantity Q_{DOM} will be demanded domestically at a price of P_{DOM}, which is above the export price P_{FOR}. Thus the firm is indeed dumping, selling more cheaply abroad than at home.

In both our numerical example and Figure 6-8, the reason the firm chooses to dump is the difference in the responsiveness of sales to price in the export and domestic markets. In Figure 6-8 we assume the firm can increase exports without cutting its price, so marginal revenue and price coincide on the export market. Domestically, by contrast, increased sales

[7] It might seem that the monopolist should set domestic sales at the level where *MC* and MR_{DOM} intersect. But remember that the monopolist is producing a total output $Q_{MONOPOLY}$; this means that the cost of producing one more unit is equal to P_{FOR}, whether that unit is destined for the foreign or domestic market. And it is the actual cost of producing one more unit that must be set equal to marginal revenue. The intersection of *MC* and MR_{DOM} is where the firm would produce *if it did not have the option of exporting*—but that is irrelevant.

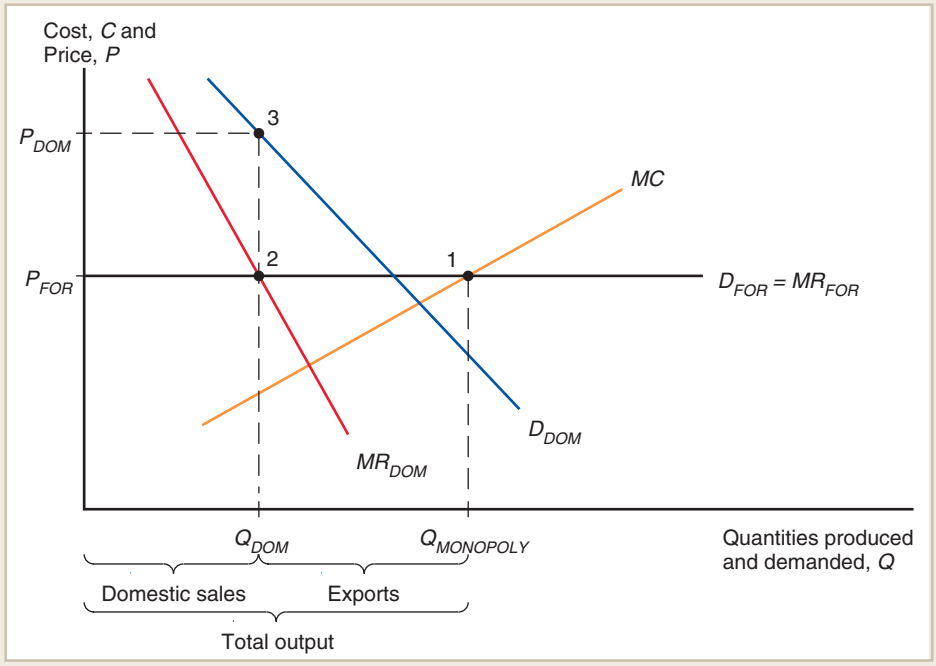


Figure 6-8

Dumping

The figure shows a monopolist that faces a demand curve D_{DOM} for domestic sales, but which can also sell as much as it likes at the export price P_{FOR}. Since an additional unit can always be sold at P_{FOR}, the firm increases output until the marginal cost equals P_{FOR}, this profit-maximizing output is shown as $Q_{MONOPOLY}$. Since the firm's marginal cost at $Q_{MONOPOLY}$ is P_{FOR}, it sells output on the domestic market up to the point where marginal revenue equals P_{FOR}; this profit-maximizing level of domestic sales is shown as Q_{DOM}. The rest of its output, $Q_{MONOPOLY} - Q_{DOM}$, is exported.

The price at which domestic consumers demand Q_{DOM} is P_{DOM}. Since $P_{DOM} > P_{FOR}$, the firm sells exports at a lower price than it charges domestic consumers.

do lower the price. This is an extreme example of the general condition for price discrimination presented in microeconomics courses: Firms will price-discriminate when sales are more price-responsive in one market than in another.[8] (In this case we have assumed export demand is infinitely price-responsive.)

Dumping is widely regarded as an unfair practice in international trade. There is no good economic justification for regarding dumping as particularly harmful, but U.S. trade law prohibits foreign firms from dumping in our market and automatically imposes tariffs when such dumping is discovered.

The situation shown in Figure 6-8 is simply an extreme version of a wider class of situations in which firms have an incentive to sell exports for a lower price than the price they charge domestic customers.

[8] The formal condition for price discrimination is that firms will charge lower prices in markets in which they face a higher *elasticity* of demand, where the elasticity is the percentage decrease in sales that results from a 1 percent increase in price. Firms will dump if they perceive a higher elasticity on export sales than on domestic sales.

Case Study

Antidumping as Protectionism

In the United States and a number of other countries, dumping is regarded as an unfair competitive practice. Firms that claim to have been injured by foreign firms who 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 "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 dumping out as a prohibited practice. For one thing, price discrimination between markets may be 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 export customers, the United States and other nations instead often try to calculate a supposed 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 universal 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 it 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."[9]

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.

Most economists consider these kinds of "antidumping" cases to have little to do with dumping in the economic sense. Nonetheless, there may have been an increase in real dumping, because of the uneven pace at which countries have opened up their

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

markets. Since 1970 trade liberalization and deregulation have opened up international competition in a number of previously sheltered industries. For example, it used to be taken for granted that telephone companies would buy their equipment from domestic manufacturers. With the breakup of AT&T in the United States and the privatization of phone companies in other countries, this is no longer the case everywhere. But in Japan and several European countries the old rules still apply. It is not surprising that the manufacturers of telephone equipment in these countries would continue to charge high prices at home while offering lower prices to customers in the United States—or at least that they would be accused of doing so.

Reciprocal Dumping

The analysis of dumping suggests that price discrimination can actually give rise to international trade. Suppose there are two monopolies, each producing the same good, one in Home and one in Foreign. To simplify the analysis, assume that these two firms have the same marginal cost. Suppose also that there are some costs of transportation between the two markets, so that if the firms charge the same price there will be no trade. In the absence of trade, each firm's monopoly would be uncontested.

If we introduce the possibility of dumping, however, trade may emerge. Each firm will limit the quantity it sells in its home market, recognizing that if it tries to sell more it will drive down the price on its existing domestic sales. If a firm can sell a little bit in the other market, however, it will add to its profits even if the price is lower than in the domestic market, because the negative effect on the price of existing sales will fall on the other firm, not on itself. So each firm has an incentive to "raid" the other market, selling a few units at a price that (net of transportation costs) is lower than the home market price but still above marginal cost.

If both firms do this, however, the result will be the emergence of trade even though there was (by assumption) no initial difference in the price of the good in the two markets, and even though there are some transportation costs. Even more peculiarly, there will be two-way trade in the same product. For example, a cement plant in country A might be shipping cement to country B while a cement plant in B is doing the reverse. The situation in which dumping leads to two-way trade in the same product is known as **reciprocal dumping.**[10]

This may seem like a strange case, and it is admittedly probably rare in international trade for exactly identical goods to be shipped in both directions at once. However, the reciprocal dumping effect probably tends to increase the volume of trade in goods that are not quite identical.

Is such peculiar and seemingly pointless trade socially desirable? The answer is ambiguous. It is obviously wasteful to ship the same good, or close substitutes, back and forth when transportation is costly. However, notice that the emergence of reciprocal dumping in our story eliminates what were initially pure monopolies, leading to some

[10] The possibility of reciprocal dumping was first noted by James Brander, "Intraindustry Trade in Identical Commodities," *Journal of International Economics* 11 (1981), pp. 1–14.

competition. The increased competition represents a benefit that may offset the waste of resources in transportation. The net effect of such peculiar trade on a nation's economic welfare is therefore uncertain.

The Theory of External Economies

In the monopolistic competition model of trade, it is presumed that the economies of scale that give rise to international trade occur at the level of the individual firm. That is, the larger any particular firm's output of a product, the lower its average cost. The inevitable result of such economies of scale at the level of the firm is imperfect competition, which in turn allows such practices as dumping.

As we pointed out early in this chapter, however, not all scale economies apply at the level of the individual firm. For a variety of reasons, it is often the case that concentrating production of an industry in one or a few locations reduces the industry's costs, even if the individual firms in the industry remain small. When economies of scale apply at the level of the industry rather than at the level of the individual firm, they are called *external economies*. The analysis of external economies goes back more than a century to the British economist Alfred Marshall, who was struck by the phenomenon of "industrial districts"—geographical concentrations of industry that could not be easily explained by natural resources. In Marshall's time, the most famous examples included such concentrations of industry as the cluster of cutlery manufacturers in Sheffield and the cluster of hosiery firms in Northampton. Modern examples of industries where there seem to be powerful external economies include the semiconductor industry, concentrated in California's famous Silicon Valley; the investment banking industry, concentrated in New York; and the entertainment industry, concentrated in Hollywood.

Marshall argued that there were three main reasons why a cluster of firms may be more efficient than an individual firm in isolation: the ability of a cluster to support **specialized suppliers;** the way that a geographically concentrated industry allows **labor market pooling;** and the way that a geographically concentrated industry helps foster **knowledge spillovers.** These same factors continue to be valid today.

Specialized Suppliers

In many industries, the production of goods and services—and to an even greater extent, the development of new products—requires the use of specialized equipment or support services; yet an individual company does not provide a large enough market for these services to keep the suppliers in business. A localized industrial cluster can solve this problem by bringing together many firms that collectively provide a large enough market to support a wide range of specialized suppliers. This phenomenon has been extensively documented in Silicon Valley: A 1994 study recounts how, as the local industry grew, "engineers left established semiconductor companies to start firms that manufactured capital goods such as diffusion ovens, step-and-repeat cameras, and testers, and materials and components such as photomasks, testing jigs, and specialized chemicals. . . . This independent equipment sector promoted the continuing formation of semiconductor firms by freeing individual producers from the expense of developing capital equipment internally and by spreading the costs of development. It also reinforced the tendency toward industrial localization, as most of these specialized inputs were not available elsewhere in the country."[11]

[11] See the book listed in Further Reading by Saxenian, p. 40.

As the quote suggests, the availability of this dense network of specialized suppliers has given high-technology firms in Silicon Valley some considerable advantages over firms elsewhere. Key inputs are cheaper and more easily available because there are many firms competing to provide them, and firms can concentrate on what they do best, contracting out other aspects of their business. For example, some Silicon Valley firms that specialize in providing highly sophisticated computer chips for particular customers have chosen to become "fabless," that is, they do not have any factories in which chips can be fabricated. Instead, they concentrate on designing the chips, then hire another firm actually to fabricate them.

A company that tried to enter the industry in another location—for example, in a country that did not have a comparable industrial cluster—would be at an immediate disadvantage because it would lack easy access to Silicon Valley's suppliers and would either have to provide them for itself or be faced with the task of trying to deal with Silicon Valley–based suppliers at long distance.

Labor Market Pooling

A second source of external economies is the way that a cluster of firms can create a pooled market for workers with highly specialized skills. Such a pooled market is to the advantage of both the producers and the workers as the producers are less likely to suffer from labor shortages, while the workers are less likely to become unemployed.

The point can best be made with a simplified example. Imagine that there are two companies that both use the same kind of specialized labor, say, two film studios that make use of experts in computer animation. Both employers are, however, uncertain about how many workers they will want to hire: If demand for its product is high, both companies will want to hire 150 workers, but if it is low, they will only want to hire 50. Suppose also that there are 200 workers with this special skill. Now compare two situations: one with both firms and all 200 workers in the same city, the other with the firms and 100 workers in two different cities. It is straightforward to show that both the workers and their employers are better off if everyone is in the same place.

First, consider the situation from the point of view of the companies. If they are in different locations, whenever one of the companies is doing well it will be confronted with a labor shortage; it will want to hire 150 workers, but only 100 will be available. If the firms are near each other, however, it is at least possible that one will be doing well when the other is doing badly, so that both firms may be able to hire as many workers as they want. So by locating near each other, the companies increase the likelihood that they will be able to take advantage of business opportunities.

From the workers' point of view, having the industry concentrated in one location is also an advantage. If the industry is divided between two cities, then whenever one of the firms has a low demand for workers the result will be unemployment; the firm will be willing to hire only 50 of the 100 workers who live nearby. But if the industry is concentrated in a single city, low labor demand from one firm will at least sometimes be offset by high demand from the other. As a result, workers will have a lower risk of unemployment.

Again, these advantages have been documented for Silicon Valley, where it is common both for companies to expand rapidly and for workers to change employers. The same study of Silicon Valley that was quoted previously notes that the concentration of firms in a single location makes it easy to switch employers, quoting one engineer as saying that "it wasn't that big a catastrophe to quit your job on Friday and have another job on Monday. . . . You didn't even necessarily have to tell your wife. You

just drove off in another direction on Monday morning."[12] This flexibility makes Silicon Valley an attractive location both for highly skilled workers and for the companies that employ them.

Knowledge Spillovers

It is by now a cliché that in the modern economy knowledge is at least as important an input as factors of production like labor, capital, and raw materials. This is especially true in highly innovative industries, where being only a few months behind the cutting edge in production techniques or product design can put a company at a major disadvantage.

But where does the specialized knowledge that is crucial to success in innovative industries come from? Companies can acquire technology through their own research and development efforts. They can also try to learn from competitors by studying their products and, in some cases, taking them apart to "reverse engineer" their design and manufacture. An important source of technical know-how, however, is the informal exchange of information and ideas that takes place at a personal level. And this kind of informal diffusion of knowledge often seems to take place most effectively when an industry is concentrated in a fairly small area, so that employees of different companies mix socially and talk freely about technical issues.

Marshall described this process memorably when he wrote that in a district with many firms in the same industry, "The mysteries of the trade become no mystery, but are as it were in the air. . . . Good work is rightly appreciated, inventions and improvements in machinery, in processes and the general organization of the business have their merits promptly discussed: If one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas."[13]

A journalist described how these knowledge spillovers worked during the rise of Silicon Valley (and also gave an excellent sense of the amount of specialized knowledge involved in the industry) as follows: "Every year there was some place, the Wagon Wheel, Chez Yvonne, Rickey's, the Roundhouse, where members of this esoteric fraternity, the young men and women of the semiconductor industry, would head after work to have a drink and gossip and trade war stories about phase jitters, phantom circuits, bubble memories, pulse trains, bounceless contacts, burst modes, leapfrog tests, p-n junctions, sleeping sickness modes, slow-death episodes, RAMs, NAKs, MOSes, PCMs, PROMs, PROM blowers, PROM blasters, and teramagnitudes. . . . "[14] This kind of informal information flow means that it is easier for companies in the Silicon Valley area to stay near the technological frontier than it is for companies elsewhere; indeed, many multinational firms have established research centers and even factories in Silicon Valley simply in order to keep up with the latest technology.

External Economies and Increasing Returns

A geographically concentrated industry is able to support specialized suppliers, provide a pooled labor market, and facilitate knowledge spillovers in a way that a geographically dispersed industry cannot. But a country cannot have a large concentration of firms in an industry unless it possesses a large industry. Thus the theory of external economies indicates that when these external economies are important, a country with a large industry

12 Saxenian, p. 35.

13 Alfred Marshall, *Principles of Economics* (London: MacMillan, 1920).

14 Tom Wolfe, quoted in Saxenian, p. 33.

will, other things being equal, be more efficient in that industry than a country with a small industry. Or to put it differently, external economies can give rise to increasing returns to scale *at the level of the national industry.*

While the details of external economies in practice are often quite subtle and complex (as the example of Silicon Valley shows), it can be useful to abstract from the details and represent external economies simply by assuming that an industry's costs are lower, the larger the industry. If we ignore possible imperfections in competition, this means that the industry will have a **forward-falling supply curve:** The larger the industry's output, the lower the price at which firms are willing to sell their output.

External Economies and International Trade

External economies, like economies of scale that are internal to firms, play an important role in international trade, but they may be quite different in their effects. In particular, external economies can cause countries to get "locked in" to undesirable patterns of specialization and can even lead to losses from international trade.

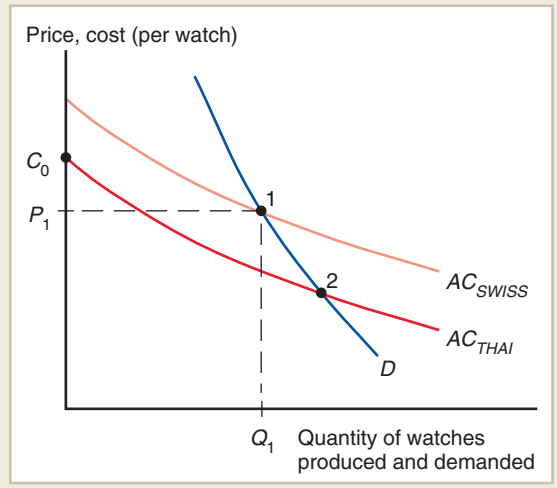
External Economies and the Pattern of Trade

When there are external economics of scale, a country that has large production in some industry will tend, other things equal, to have low costs of producing that good. This gives rise to an obvious circularity, since a country that can produce a good cheaply will also therefore tend to produce a lot of that good. Strong external economies tend to confirm existing patterns of interindustry trade, whatever their original sources: Countries that start out as large producers in certain industries, for whatever reason, tend to remain large producers. They may do so even if some other country could potentially produce the goods more cheaply.

Figure 6-9 illustrates this point. We show the cost of producing a watch as a function of the number of watches produced annually. Two countries are shown: "Switzerland" and "Thailand." The Swiss cost of producing a watch is shown as AC_{SWISS}; the Thai cost as

Figure 6-9
External Economics and Specialization

The average cost curve for Thailand, AC_{THAI} lies below the average cost curve for Switzerland, AC_{SWISS}. Thus Thailand could potentially supply the world market more cheaply than Switzerland. If the Swiss industry gets established first, however, it may be able to sell watches at the price P_1 which is below the cost C_0 that an individual Thai firm would face if it began production on its own. So a pattern of specialization established by historical accident may persist even when new producers could potentially have lower costs.



AC_{THAI}. D represents the world demand for watches, which we assume can be satisfied either by Switzerland or by Thailand.

Suppose that the economies of scale in watch production are entirely external to firms, and that since there are no economies of scale at the level of the firm the watch industry in each country consists of many small perfectly competitive firms. Competition therefore drives the price of watches down to its average cost.

We assume that the Thai cost curve lies below the Swiss curve, say because Thai wages are lower than Swiss. This means that at any given level of production, Thailand could manufacture watches more cheaply than Switzerland. One might hope that this would always imply that Thailand will in fact supply the world market. Unfortunately, this need not be the case. Suppose that Switzerland, for historical reasons, establishes its watch industry first. Then initially world watch equilibrium will be established at point 1 in Figure 6-9, with Swiss production of Q_1 units per year and a price of P_1. Now introduce the possibility of Thai production. If Thailand could take over the world market, the equilibrium would move to point 2. However, if there is no initial Thai production ($Q = 0$) any individual Thai firm considering manufacture of watches will face a cost of production of C_0. As we have drawn it, this cost is above the price at which the established Swiss industry can produce watches. So although the Thai industry could potentially make watches more cheaply than Switzerland, Switzerland's head start enables it to hold on to the industry.

As this example shows, external economies potentially give a strong role to historical accident in determining who produces what, and may allow established patterns of specialization to persist even when they run counter to comparative advantage.

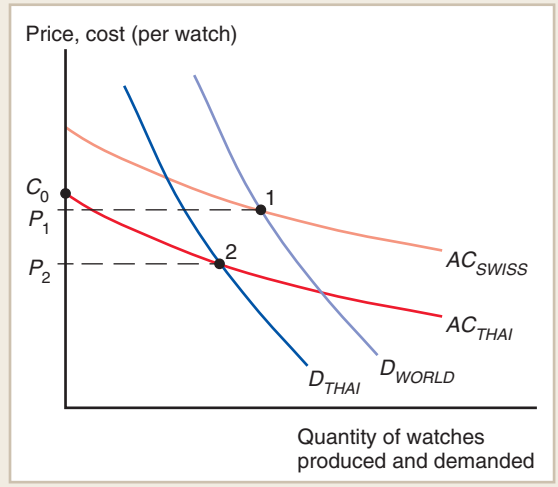
Trade and Welfare with External Economies

Trade based on external economies has more ambiguous effects on national welfare than either trade based on comparative advantage or trade based on economies of scale at the level of the firm. There may be gains to the world economy from concentrating production in particular industries to realize external economies. On the other hand, there is no guarantee that the right country will produce a good subject to external economies, and it is possible that trade based on external economies may actually leave a country worse off than it would have been in the absence of trade.

Figure 6-10

External Economics and Losses from Trade

When there are external economies, trade can potentially leave a country worse off than it would be in the absence of trade. In this example, Thailand imports watches from Switzerland, which is able to supply the world market (D_{WORLD}) at a price (P_1) low enough to block entry by Thai producers who must initially produce the watches at cost C_0. Yet if Thailand were to block all trade in watches, it would be able to supply its domestic market (D_{THAI}) at the lower price P_2.



An example of how a country can actually be worse off with trade than without is shown in Figure 6-10. In this example, as before, we imagine that Thailand and Switzerland could both manufacture watches, that Thailand could make them more cheaply, but that Switzerland has gotten there first. D_{WORLD} is the world demand for watches, and, given that Switzerland produces the watches, the equilibrium is at point 1. However, we now add to the figure the Thai demand for watches, D_{THAI}. If no trade in watches were allowed and Thailand were forced to be self-sufficient, then the Thai equilibrium would be at point 2. Because of its lower average cost curve, the price of Thai-made watches at point 2, P_2, is actually lower than the price of Swiss-made watches at point 1, P_1.

We have shown a situation in which the price of a good that Thailand imports would actually be lower if there were no trade and the country were forced to produce the good for itself. Clearly in this situation trade leaves the country worse off than it would be in the absence of trade.

There is an incentive in this case for Thailand to protect its potential watch industry from foreign competition. Before concluding that this justifies protectionism, however, we should note that in practice identifying cases like that in Figure 6-10 is far from easy. Indeed, as we will emphasize in Chapters 10 and 11, the difficulty of identifying external economies in practice is one of the main arguments against activist government policies toward trade.

It is also worth pointing out that while external economies can sometimes lead to disadvantageous patterns of specialization and trade, it is still to the benefit of the *world* economy to take advantage of the gains from concentrating industries. Canada might be better off if Silicon Valley were near Toronto instead of San Francisco; Germany might be better off if the City (London's financial district, which, along with Wall Street, dominates world financial markets) could be moved to Frankfurt. The world as a whole is, however, more efficient and thus richer because international trade allows nations to specialize in different industries and thus reap the gains from external economies as well as the gains from comparative advantage.

Dynamic Increasing Returns

Some of the most important external economies probably arise from the accumulation of knowledge. When an individual firm improves its products or production techniques through experience, other firms are likely to imitate the firm and benefit from its knowledge. This spillover of knowledge gives rise to a situation in which the production costs of individual firms fall as the industry as a whole accumulates experience.

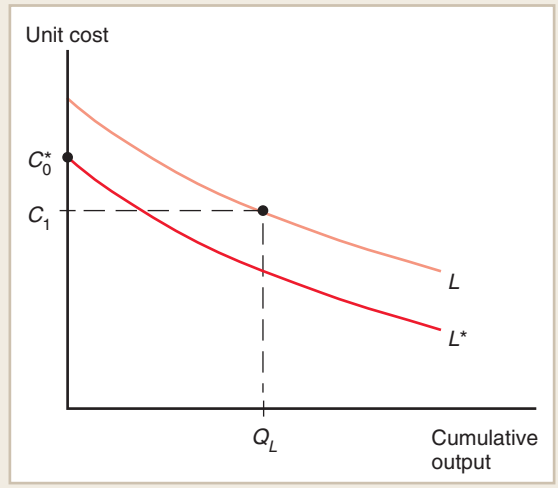
Notice that external economies arising from the accumulation of knowledge differ somewhat from the external economies considered so far, in which industry costs depend on current output. In this alternative situation industry costs depend on experience, usually measured by the cumulative output of the industry to date. For example, the cost of producing a ton of steel might depend negatively on the total number of tons of steel produced by a country since the industry began. This kind of relationship is often summarized by a **learning curve** that relates unit cost to cumulative output. Such learning curves are illustrated in Figure 6-11. They are downward sloping because of the effect of the experience gained through production on costs. When costs fall with cumulative production over time, rather than with the current rate of production, this is referred to as a case of **dynamic increasing returns.**

Like ordinary external economies, dynamic external economies can lock in an initial advantage or head start in an industry. In Figure 6-11, the learning curve L is that of a country that pioneered an industry, while L^* is that of another country that has lower input

Figure 6-11

The Learning Curve

The learning curve shows that unit cost is lower the greater the cumulative output of a country's industry to date. A country that has extensive experience in an industry (L) may have lower unit cost than another country with little or no experience, even if the second country's learning curve (L^*) is lower, for example, because of lower wages.



costs—say, lower wages—but less production experience. Provided that the first country has a sufficiently large head start, the potentially lower costs of the second country may not allow it to enter the market. For example, suppose the first country has a cumulative output of Q_L units, giving it a unit cost of C_1 while the second country has never produced the good. Then the second country will have an initial start-up cost C^*_0 that is higher than the current unit cost, C_1, of the established industry.

Dynamic scale economics, like external economics at a point in time, potentially justify protectionism. Suppose that a country could have low enough costs to produce a good for export if it had more production experience, but that given the current lack of experience the good cannot be produced competitively. Such a country might increase its long-term welfare either by encouraging the production of the good by a subsidy or by protecting it from foreign competition until the industry could stand on its own feet. The argument for temporary protection of industries to enable them to gain experience is known as the **infant industry argument** and has played an important role in debates over the role of trade policy in economic development. We will discuss the infant industry argument at greater length in Chapter 10, but for now we simply note that situations like that illustrated in Figure 6-11 are just as hard to identify in practice as in those involving nondynamic increasing returns.

Interregional Trade and Economic Geography

External economies play an important role in shaping the pattern of international trade, but they are even more decisive in shaping the pattern of **interregional trade**—trade that takes place between regions *within* countries.

To understand the role of external economies in interregional trade, we first need to discuss the nature of regional economics—that is, how the economies of regions within a nation fit into the national economy. Studies of the location of U.S. industries suggest that more than 60 percent of U.S. workers are employed by industries whose output is nontradable even within the United States—that is, which must be supplied locally. Table 6-4 shows some examples of tradable and nontradable industries. Thus, motion pictures made in Hollywood are shown across the country, and indeed around the world, but newspapers are mainly read in their home cities. Wall Street trades stocks and makes deals for clients

Tinseltown Economics

What is the United States' most important export sector? The answer depends to some extent on definitions; some people will tell you that it is agriculture, others that it is aircraft. By any measure, however, one of the biggest exporters in the United States is the entertainment sector, movies in particular. In 2000, rental fees generated by exports of films and tape were $8.9 billion—more than domestic box office receipts. American films dominated ticket sales in much of the world, accounting for 82 percent of box office receipts in Germany, 65 percent in Japan, and 58 percent in France.

Why is the United States the world's dominant exporter of entertainment? There are important advantages arising from the sheer size of the American market. A film aimed primarily at the French or Italian markets, which are far smaller than that of the United States, cannot justify the huge budgets of many American films. Thus films from these countries are typically dramas or comedies whose appeal fails to survive dubbing or subtitles. Meanwhile, American films can transcend the language barrier with lavish productions and spectacular special effects.

But an important part of the American dominance in the industry also comes from the external economies created by the immense concentration of entertainment firms in Hollywood. Hollywood clearly generates two of Marshall's types of external economies: specialized suppliers and labor market pooling. While the final product is provided by movie studios and television networks, these in turn draw on a complex web of independent producers, casting and talent agencies, legal firms, special effects experts, and so on. And the need for labor market pooling is obvious to anyone who has watched the credits at the end of a movie: Each production requires a huge but temporary army that includes not just cameramen and makeup artists but musicians, stunt men and women, and mysterious occupations like gaffers and grips (and—oh yes—actors and actresses). Whether it also generates the third kind of external economies—knowledge spillovers—is less certain. After all, as the author Nathaniel West once remarked, the key to understanding the movie business is to realize that "nobody knows anything." Still, if there is any knowledge to spill over, surely it does so better in the intense social environment of Hollywood than it could anywhere else.

An indication of the force of Hollywood's external economies has been its persistent ability to draw talent from outside the United States. From Garbo and von Sternberg to Arnold Schwarzenegger and Paul Verhoeven, "American" films have often been made by ambitious foreigners who moved to Hollywood—and in the end reached a larger audience even in their original nations than they could have if they had remained at home.

Is Hollywood unique? No, similar forces have led to the emergence of several other entertainment complexes. In India, whose film market has been protected from American domination partly by government policy and partly by cultural differences, a movie-making cluster known as "Bollywood" has emerged in Bombay. A substantial film industry catering to Chinese speakers has emerged in Hong Kong. And a specialty industry producing Spanish-language television programs for all of Latin America, focusing on so-called *telenovelas,* long-running soap operas, has emerged in Caracas, Venezuela. This last entertainment complex has discovered some unexpected export markets: Television viewers in Russia, it turns out, identify more readily with the characters in Latin American soaps than with those in U.S. productions.

TABLE 6-4 Some Examples of Tradable and Nontradable Industries

Tradable industries	Nontradable industries
Motion pictures	Newspaper publishers
Securities, commodities, etc.	Savings institutions
Scientific research	Veterinary services

Source: J. Bradford Jensen and Lori. G. Kletzer, "Tradable Services: Understanding the Scope and Impact of Services Outsourcing," Peterson Institute of International Economics Working Paper WP 05-9, September 2005.

across America, but savings banks mainly serve local depositors. Scientists at the National Institutes for Health develop medical knowledge that is applied across the whole country, but the veterinarian who figures out why your pet is sick has to be near your home.

As you might expect, the share of nontradable industries in employment is pretty much the same across the United States. For example, restaurants employ about 5 percent of the workforce in every major U.S. city. On the other hand, tradable industries vary greatly in importance across regions. Manhattan accounts for only 1.9% percent of America's total employment, but it accounts for 24.5% of those employed in trading stocks and bonds and 14.7% of employment in the advertising industry.

But what determines the location of tradable industries? In some cases, natural resources play a key role—for example, Houston is a center for the oil industry because east Texas is where the oil is. However, factors of production such as labor and capital play a less decisive role in interregional trade than in international trade, for the simple reason that such factors are highly mobile within countries. (We'll discuss the role of factor movements between countries in Chapter 7.) As a result, factors tend to move to where the industries are rather than the other way around. For example, California's Silicon Valley, near San Francisco, has a very highly educated labor force, with a high concentration of engineers and computer experts; that's not because California trains lots of engineers, it's because engineers move to Silicon Valley to take jobs with the region's high-tech industry.

Resources, then, play a secondary role in interregional trade. What largely drives specialization and trade, instead, is external economies. Why, for example, are so many advertising agencies located in New York? The answer is, because so many *other* advertising agencies are located in New York. As one recent study put it, "Information sharing and information diffusion are critical to a team and an agency's success. . . . In cities like New York, agencies group in neighborhood clusters. Clusters promote localized networking, to enhance creativity; agencies share information and ideas and in doing this face-to-face contact is critical."[15] In fact, the evidence suggests that the external economies that support the advertising business are *very* localized: To reap the benefits of information spillovers, ad agencies need to be located within about 300 yards of each other!

But if external economies are the main reason for regional specialization and interregional trade, what explains how a particular region develops the external economies that support an industry? The answer, in general, is that accidents of history play a crucial role. A century and a half ago, New York was America's most important port city because it had access to the Great Lakes via the Erie Canal. That led to New York's becoming America's financial center; it remains America's financial center today thanks to the external economies the financial industry creates for itself. Los Angeles became the center of the early film industry when films were shot outdoors and needed good weather; it remains the center of the film industry today, even though many films are shot indoors or on location, because of the externalities described in the box on p. 147.

A question you might ask is whether the forces driving interregional trade are really all that different from those driving international trade. The answer is that they are not, especially when one looks at trade between closely integrated national economies, such as those of western Europe. Indeed, London plays a role as Europe's financial capital similar to that played by New York as America's financial capital. In recent years, there has been a growing movement among economists to model interregional and international trade, as well as such phenomena as the rise of cities, as different aspects of the same phenomenon—economic interaction across space. Such an approach is often referred to as **economic geography.**

[15] J. Vernon Henderson, "What Makes Big Cities Tick? A Look at New York," mimeo, Brown University, 2004.

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 normally lead to a breakdown of perfect competition, so that trade in the presence of economies of scale must be analyzed using models of imperfect competition. Two important models of this kind are the monopolistic competition model and the dumping model. A third model, that of external economies, is consistent with perfect competition.
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 larger scale and thus lower average cost, than a small market.
4. International trade allows creation of an integrated market that is larger than any one country's market, and thus makes it possible simultaneously to offer consumers a greater variety of products and lower prices.
5. In the monopolistic competition model, trade may be divided into two kinds. Two-way trade in differentiated products within an industry is called intraindustry trade; trade that exchanges the products of one industry for the products of another is called interindustry trade. Intraindustry trade reflects economies of scale; interindustry trade reflects comparative advantage. Intraindustry trade does not generate the same strong effects on income distribution as interindustry trade.
6. Dumping occurs when a monopolistic firm charges a lower price on exports than it charges domestically. It is a profit-maximizing strategy when export sales are more price-responsive than domestic sales, and when firms can effectively segment markets, that is, prevent domestic customers from buying goods intended for export markets. Reciprocal dumping occurs when two monopolistic firms dump into each others' home markets; such reciprocal dumping can be a cause of international trade.
7. External economies are economies of scale that occur at the level of the industry instead of the firm. They give an important role to history and accident in determining the pattern of international trade. When external economies are important, a country starting with a large industry may retain that advantage even if another country could potentially produce the same goods more cheaply. When external economies are important, countries can conceivably lose from trade.

KEY TERMS

average cost, p. 119
dumping, p. 135
dynamic increasing returns, p. 145
economic geography, p. 148
external economies of scale, p. 116
forward-falling supply curve, p. 143
imperfect competition, p. 117
infant industry argument, p. 146
interindustry trade, p. 131
internal economies of scale, p. 116
interregional trade p. 146
intraindustry trade, p. 131
knowledge spillovers, p. 140
labor market pooling, p. 140
learning curve, p. 145
marginal cost, p. 119
marginal revenue, p. 118
monopolistic competition, p. 120
oligopoly, p. 120
price discrimination, p. 135
pure monopoly, p. 117
reciprocal dumping, p. 139
specialized suppliers, p. 140

PROBLEMS

myeconlab

1. For each of the following examples, explain whether it is a case of external or internal economies of scale:
   a. Most musical wind instruments in the United States are produced by more than a dozen factories in Elkhart, Indiana.
   b. All Hondas sold in the United States are either imported or produced in Marysville, Ohio.
   c. All airframes for Airbus, Europe's only producer of large aircraft, are assembled in Toulouse, France.
   d. Hartford, Connecticut, is the insurance capital of the northeastern United States.
2. In perfect competition, firms set price equal to marginal cost. Why isn't this possible when there are internal economies of scale?
3. It is often argued that the existence of increasing returns is a source of conflict between countries, since each country is better off if it can increase its production in those industries characterized by economies of scale. Evaluate this view in terms of both the monopolistic competition and the external economy models.
4. Suppose the two countries we considered in the numerical example on pages 127–129 were to integrate their automobile market with a third country with 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.
5. 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.[16]
   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 in Europe combined? What will be the new equilibrium price of automobiles?
   d. Why are prices in the United States different in (c) than in (b)? Are consumers better off with free trade? In what ways?
6. Give two examples of products that are traded on international markets for which there are dynamic increasing returns. In each of your examples, show how innovation and learning-by-doing are important to the dynamic increasing returns in the industry.
7. Evaluate the relative importance of economies of scale and comparative advantage in causing the following:
   a. Most of the world's aluminum is smelted in Norway or Canada.
   b. Half of the world's large jet aircraft are assembled in Seattle.
   c. Most semiconductors are manufactured in either the United States or Japan.
   d. Most Scotch whiskey comes from Scotland.
   e. Much of the world's best wine comes from France.

[16] This question has been modified. Previous printings of the textbook stated that "$P = 8{,}000 + (150/n)$," which has changed to "$P = 17{,}000 + (150/n)$."

8. There are some shops in Japan that sell *Japanese* goods imported back from the United States at a discount over the prices charged by other Japanese shops. How is this possible?
9. Consider a situation similar to that in Figure 6-9, in which two countries that can produce a good are subject to forward-falling supply curves. In this case, however, suppose that the two countries have the same costs, so that their supply curves are identical.
   a. What would you expect to be the pattern of international specialization and trade? What would determine who produces the good?
   b. What are the *benefits* of international trade in this case? Do they accrue only to the country that gets the industry?
10. It is fairly common for an industrial cluster to break up and for production to move to locations with lower wages when the technology of the industry is no longer rapidly improving—when it is no longer essential to have the absolutely most modern machinery, when the need for highly skilled workers has declined, and when being at the cutting edge of innovation conveys only a small advantage. Explain this tendency of industrial clusters to break up in terms of the theory of external economies.
11. Which of the following goods or services would be most likely to be subject to (1) external economies of scale and (2) dynamic increasing returns? Explain your answers.
   a. Software tech-support services
   b. Production of asphalt or concrete
   c. Motion pictures
   d. Cancer research
   e. Timber harvesting

FURTHER READING

Frank Graham. "Some Aspects of Protection Further Considered." *Quarterly Journal of Economics* 37 (1923), pp. 199–227. An early warning that international trade may be harmful in the presence of external economies of scale.

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.

Staffan Burenstam Linder. *An Essay on Trade and Transformation.* New York: John Wiley and Sons, 1961. An early and influential statement of the view that trade in manufactures among advanced countries mainly reflects forces other than comparative advantage.

Michael Porter. *The Competitive Advantage of Nations.* New York: Free Press, 1990. A best-selling book that explains national export success as the result of self-reinforcing industrial clusters, that is, external economies.

Annalee Saxenian. *Regional Advantage.* Cambridge: Harvard University Press, 1994. A fascinating comparison of two high-technology industrial districts, California's Silicon Valley and Boston's Route 128.

APPENDIX TO CHAPTER 6

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

$$Q = A - B \times P, \tag{6A-1}$$

its marginal revenue is

$$MR = P - (1/B) \times Q. \tag{6A-2}$$

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 (6A-1) we get

$$P = (A/B) - (1/B) \times Q. \tag{6A-3}$$

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

$$R = P \times Q = [(A/B) - (1/B) \times Q] \times Q. \tag{6A-4}$$

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

$$\begin{aligned} R' = P' \times Q' &= [(A/B) - (1/B) \times (Q + dQ)] \times (Q + dQ) \\ &= [(A/B) - (1/B) \times Q] \times Q + [(A/B) - (1/B) \times Q] \times dQ \\ &\quad - (1/B) \times Q \times dQ - (1/B) \times (dQ)^2. \end{aligned} \tag{6A-5}$$

Equation (6A-5) can be simplified by substitution in from (6A-1) and (6A-4) to get

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

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 (6A-6) can be ignored. This gives us the result that the *change* in revenue from a small change in sales is

$$R' - R = [P - (1/B) \times Q] \times dQ. \tag{6A-7}$$

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 (6A-2).