FEDERAL ANTITRUST POLICY

THE LAW OF COMPETITION AND ITS PRACTICE

Second Edition

By

HERBERT HOVENKAMP
Ben V. & Dorothy Willie Professor
University of Iowa, College of Law

HORNBOOK SERIES®

ST. PAUL, MINN., 1999

Part I

FOUNDATIONS: POLICY AND MEASUREMENT

Chapter 1

THE BASIC ECONOMICS OF ANTITRUST

Table of Sections

§ 1.1 Price Theory: Economic Behavior and Perfect Competition

Those who make antitrust policy are consumers, not usually creators, of economic theory. Further, antitrust policy makers are quite stodgy about adopting new theory. The economics applied in antitrust decision making is quite conventional, "applied" economics. The economics literature as a whole is more technical, more venturesome and speculative, much more stylized, and at the margins much more controversial than most of the economics that is applied by the antitrust policy maker.[1] What follows is a brief presentation of relatively orthodox economics that forms the basis of federal antitrust policy.

§ 1.1

1. For example, consider the literature on game theory that now forms the center of industrial organization analysis in economics departments, but is barely beginning to make inroads in applied antitrust economics. Further, the game theory being applied in antitrust is simple and quite uncontroversial. See 1 Handbook of Industrial Organization, chs. 5–7 (R. Schmalensee & R. Willig, eds. 1989). But see S. Peltzman, The Handbook of Industrial Organization: a Review Article, 99 J.Pol.Econ. 201 (1991), arguing that very little in game theory offers useful predictions; *pro tanto*, it is of little use to the policy maker. Accord T. Muris, Economics and Antitrust, 5 Geo. Mason L. Rev. 303 (1997).

Market economies are dedicated to the principle that in the first instance people are responsible for their own welfare. Further, they are best off if they can make voluntary exchanges of goods and services in competitive markets.[2] If all exchanges are voluntary, each person will continue to exchange goods and services until she can make herself no better off by an exchange that is voluntary for both parties to the transaction. If all exchanges occur at competitive prices, society as a whole is wealthier than if some occur at a higher or lower price. An important goal of antitrust law—arguably its only goal—is to ensure that markets are competitive.

1.1a. The Perfectly Competitive Market

A competitive market is one in which 1) every good is priced at the cost of producing it, giving the producers and sellers only enough profit to maintain investment in the industry; and 2) every person willing to pay this price will be able to buy it.

Most customers prefer to purchase things at the lowest possible price—even, if possible, at less than the cost of producing them. By contrast, sellers prefer to sell at a price that will give them the highest possible profits. As a result competition is not an absolutely natural state of affairs; both buyers and sellers must be forced to compete.

The conditions most conducive to competition, and which obtain perfectly in an economic model of "perfect competition," are: 1) All sellers make an absolutely homogenous product, so that customers are indifferent as to which seller they purchase from, provided that the price is the same; 2) each seller in the market is so small in proportion to the entire market that the seller's increase or decrease in output, or even its exit from the market, will not affect the decisions of other sellers in that market; 3) all resources are completely mobile, or alternatively, all sellers have the same access to needed inputs; 4) all participants in the market have good knowledge about price, output and other information about the market. As a general rule, the closer a market comes to fulfilling these conditions, the more competitively it will perform.

The perfect competition model generally assumes "constant returns to scale"—that is, that costs of production per unit remain constant at all practical rates of output. As we shall see in § 1.4, the presence of substantial economies of scale—that is, of per unit costs that decrease as output increases—can undermine the perfect competition model, particularly if a firm must acquire a large market share in order to take advantage of these scale economies.

The most important rule governing price is the law of supply and demand. Price setting in any market is a function of the relationship between the amount of a product available and the amount that consumers, at the margin, are willing to pay. If the supply is not infinite, the market allocates goods to customers based on their individual willingness to pay. For example, if all the world's steel mills produced only 1000 pounds of steel per year, customers would likely bid a very high price for the steel, which would naturally be sold to the highest bidder. The price would be determined by the marginal customer's willingness to pay—that is, by the amount that some buyer would be willing to pay for the 1000th pound. Perhaps orthodontists, who put one half ounce of steel in a set of $800 braces, would be willing to buy all the steel at $3000 per pound. In that case no steel would be sold at a lower price. If the supply of steel increased 1000–fold, however, there might be far more steel than orthodontists could use at a price of $3000 per pound. The price of steel would drop so that the market could take in additional customers who place a high value on steel but are not willing to pay $3000 per pound.

2. For a normative defense of the free market, see R. Posner, The Economics of Justice (1981). The discussion of price theory that appears in this chapter is very spare, and some may be frustrated by the brevity, the lack of mathematical proof, or the paucity of examples. Those persons are referred to any modern text on microeconomics. Good current examples are J. M. Perloff, Microeconomics (1999); R. S. Pindyck & D.L. Rubinfeld, Microeconomics (4th ed. 1998). A classic and quite technical text is G. Stigler, The Theory of Price (3d ed. 1966).

As more and more steel is produced, the market price must drop further in order to reach customers who have lower "reservation" prices. A reservation price is the highest amount that a consumer is willing to pay for a product. As the price of steel drops those customers with very high reservation prices, such as the orthodontists, can also buy steel at the lower price. In the perfect competition model all sales tend to be made at the same price, even though different groups of consumers have vastly different reservation prices. If the seller attempted to charge orthodontists $3000 per pound but automakers $3 per pound, the seller's plan would be frustrated by "arbitrage." That is, automakers would buy steel at $3.00 per pound and resell some steel to orthodontists at a price higher than $3.00 per pound but lower than $3000 per pound. If all buyers have complete information about the market, all of them will pay the same price, regardless of their reservation prices. When a market reaches this condition, it is said to be in "equilibrium."[3]

Assume that the market contains 100 sellers of steel. Each seller wants to make as much money as possible, and every buyer (regardless of his reservation price) wants to purchase steel at the lowest possible price. How much steel will be produced in the market and what will be its price?

Figure 1 illustrates how a perfectly competitive market arrives at equilibrium, or the point at which supply and demand are perfectly balanced and will not change unless the market is disturbed. The figure illustrates the market demand curve (D) and the market supply curve (S) for a single product. Since both price and output are generally positive numbers, it is common to display only the upper right quadrant of the standard two-axis graph. The vertical axis represents price, which increases from 0 as one moves upward. The horizontal axis represents output (or quantity), which increases from 0 as one moves from the origin to the right.

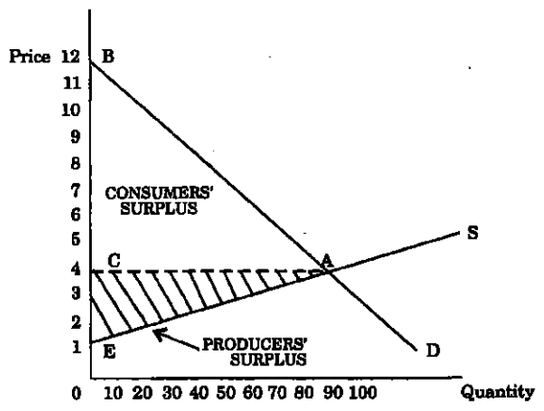

Figure 1

[608a]

The graph shows that at low levels of quantity, or output, the market price is quite high. Since few units are being produced, the good is sold only to customers who have very high reservation prices. Sellers will be earning enormous profits on their output. Profit, which is revenue (or price times quantity sold) minus cost, is measured by the vertical distance between the supply curve and the demand curve at any point. The supply curve itself includes "competitive" or "normal" profits. Any vertical distance between the supply curve and the demand curve is referred to as "economic" or "monopoly" profits. These are profits in excess of those earned by a competitive industry, and in excess of the amount needed to maintain investment in the industry.

If profits per unit of output are extremely high, as they are when output is very low, two things will happen. First, existing sellers will be encouraged by the very high profits to increase their output. Suppose current output is twenty units, the cost of production is approximately $2.00, but the price is on the order of $10.00. Each additional unit that the firm produces will give it economic profits of $8.00. Secondly, and for the same reason, new firms will come into the market. People with money to invest invariably seek opportunities where the expected return is highest.

The figure shows an upward sloping supply curve. A horizontal supply curve would imply

3. In real world markets, however, price discrimination, or obtaining higher profits from one set of customers than from another set, is both possible and common. See §§ 14.1–14.3.

that the costs of producing additional units are the same at all levels of market output. However, this is not always the case. As firms increase their output, the cost of producing the final units of output may rise. The new production must make use of increasingly marginal (less attractive) resources. The first units of steel, for example, will be produced from the iron ore that is the cheapest to obtain and refine. As output increases, however, these firms must turn to more marginal ore. Likewise, as new steel producers enter the market they will buy up the best remaining ore reserves, and firms that come in later will have to take more marginal reserves. As the market grows, increasingly marginal materials will be used and the cost of producing steel will tend to rise.[4]

As output increases, the market price will fall as customers with lower and lower reservation prices must be drawn in. The market will finally stabilize at point A. At any point on the supply curve to the left of A, an increase in output of one unit will generate positive economic profits—that is, more in revenue than the cost of producing that unit. At least one firm will increase its output or at least one new firm will enter the market and start producing. This process will continue until the supply curve and demand curve intersect.

By contrast, if production is at some point on the supply curve to the right of A, then at least some steel is being produced at less than the price than can be obtained for it. In that case the least efficient firms will exit from the market or some firms will close down their least efficient mines and plants or reduce their output until the quantity supplied falls back to the intersection with the demand curve at A. The market constantly moves toward this "equilibrium."

As noted above, in a competitive market all buyers pay the market price, even if their individual reservation prices are higher. The difference between the buyers' reservation prices and the price they actually pay is called "consumers' surplus." The size of the consumers' surplus in Figure 1 is represented by triangle ABC. A competitive market tends to maximize the size of the consumers' surplus: the consumers' surplus cannot be larger than ABC without at least one sale being unprofitable.

Some firms in the market are likely to have lower costs than others. They may have the richest veins of ore or the lowest energy, labor, or distribution costs. Cross-hatched triangle ACE represents "producers' surplus:" the difference between total revenue at the competitive price and the sum of the producers' costs. Only at the margin does a firm earn zero profits. Such a marginal firm is the one with the highest costs that is still capable of earning a competitive rate of return when the product is sold at a competitive price. If the market shifts in a way that is unfavorable to sellers, this marginal firm is likely to be the first, or one of the first, to go out of business.

The supply and demand curves in Figure 1 can assume an infinite variety of shapes. The figure shows them as straight lines, suggesting that the quantity demanded increases at a uniform rate as price falls, and that production costs rise at a uniform rate as output increases. But in most markets the two lines are nonlinear, and may often be quite irregular. Drawing them as straight lines is a useful analytic device, however, that often does not affect analysis.

The relationships expressed by the supply and demand curves can be quantified and expressed in formulas. One formula, for *price elasticity of demand*, is simply a short-hand expression for the relationship between a particular change in the price of a product and the corresponding change in demand for it. That formula is:[5]

4. If these costs differences result from new entry by additional firms, they are usually diagrammed by a supply curve that shifts to reflect higher costs. If they are the increasing costs of a single firm or group of firms, they are generally diagrammed by a curve that has an upward slope, as in Figure 1.

5. The importance of dividing δQ by Q, and δP by P in the formula is to ensure that we are talking about percentage changes, which can be expressed in any possible units of measure. That is, we might wish to express elasticity of demand simply as $\delta Q/\delta P$; but in that case the ratio would appear to change if we changed the unit in which Q is measured from, say, gallons to quarts.

$$\varepsilon = \frac{\delta Q}{Q} \div \frac{\delta P}{P^5}$$

or alternatively,

$$\varepsilon = \frac{\delta Q}{\delta P} \cdot \frac{P}{Q}$$

in which δQ and δP are equal to changes in quantity demanded and market clearing price, respectively, and Q and P are the base quantity and price from which the changes took place. Since quantity and price change in opposite directions (quantity demanded goes up as price goes down) this number is negative. As a matter of convention, however, it is common to take the absolute value, or drop the negative sign.

If at an output of 200 the market-clearing price per unit is $100, and at an output of 240 the market-clearing price drops to $90, we can compute price elasticity of demand as follows: change in output = 40; change in price = 10. The elasticity of demand equals:

$$\frac{40}{10} \cdot \frac{100}{200} = \frac{4000}{2000} = 2.$$

A simpler way of describing price elasticity of demand is that it is the relationship between the percentage change in quantity of a good demanded when the price of the good changes by a certain percentage. In the above example, a 10% drop in price elicits a 20% increase in market demand, yielding an elasticity of demand of 20%/10%, or 2.

The elasticity of demand along any curve is different than the slope of the curve, which in the case of a straight line is the ratio of the vertical axis to the horizontal axis. While the slope of a linear curve is the same at all points, the price elasticity of demand represented by a straight line demand curve is different at all points. If a demand curve stretching from the price axis to the output axis is a straight line, the elasticity of demand will be one at the line's midpoint, higher than one at all points above the midpoint, and lower than one at all points below. Whenever the elasticity of demand in a market is greater than one, we term the demand "elastic." In that case a price increase of X% will yield a decrease in quantity demanded of greater than X%. When the elasticity of demand is less than one we term the demand "inelastic." In that case a price increase of X% yields a decrease in quantity of less than X%. As you might guess, a seller would prefer to face an inelastic rather than an elastic demand: if demand is very inelastic, a relatively large price increase will yield a relatively small decrease in demand.

Elasticity of *supply* is a relationship between changes in the price of a product and the amount produced. As the price of a product rises, more of it will be produced because existing firms will increase their output or new firms will enter the market and start producing. The elasticity of supply is measured by the percentage change in the amount supplied that results from a certain percentage change in price. For example, if a 10% price increase yields a 30% increase in supply, the elasticity of supply in the market is 3. If a 30% price increase yields a 15% supply increase, the market's elasticity of supply is .5. Elasticity of supply is a positive number.

For antitrust policy one must consider not only the absolute elasticity of supply, but also the amount of time it takes for supply to increase in response to a price increase. Suppose that the elasticity of supply in a market is 3, which is very high. If price goes up by 10%, the quantity supplied to the market will increase by 30%. But suppose that the construction of the additional plants that account for the 30% supply increase takes 10 years. A seller attempting to raise its price to a monopoly level will eventually lose sales to this increased output by competitors. But during the ten year construction period the seller will earn monopoly profits. Further, the expense and time required to build a competing plant may enable the incumbent to engage in certain "strategic" behavior. For example, if prospective competitors know that the incumbent has substantial excess capacity and can increase output and drop price at will, the large investment and long wait for an uncertain return may look unprofitable.[6]

Time can also be a factor in antitrust analysis of elasticity of demand. Often customers facing a price increase can switch to a different seller more quickly than suppliers can expand output or enter the market, but this is not always the case. For example, customers may be constrained by long-term contracts, or their technology may tie them to a given supplier or group of suppliers. For example, an electric utility that uses uranium for its power plant might wish to switch to coal if the uranium market is cartelized. However, changing over is both costly and time-consuming. Eventually, when the nuclear plant wears out, the utility may switch to coal if the uranium cartel is still in existence. In its *Kodak* decision, the Supreme Court spoke of "locked-in" customers who have a large investment in a durable piece of equipment such as a photocopier, and must thus buy its specially designed replacement parts until the machine wears out or becomes obsolete.[7] Economists generally speak of these time factors by distinguishing between "long-run" and "short-run" elasticities of supply and demand. The "long-run" elasticity of supply is generally said to be higher than the "short-run" elasticity. The same is true of elasticity of demand.

The importance of time in antitrust analysis results from the fact that the policy maker is necessarily concerned with *short*-run dislocations in the market. We could presume that all markets will eventually become competitive, but antitrust is concerned with ensuring that this occurs sooner rather than later. The concern is not unique to antitrust. For example, we would not need contract law in competitive markets if our only concern was with the long run. Firms who break their contracts would be shunned by buyers and sellers who have other alternatives. Likewise, in the long run all of us will be dead. But that fact does not undermine the state's concern to protect us from murderers or see to it that we are provided with nutrition and health care.

The previous discussion of the relationship between supply and demand assumes that the market is unaffected by changes imposed from outside. If relative consumer income rises or falls, new technology makes a product obsolete or the country goes to war, however, demand for any good may rise or fall regardless of available supply or costs of production. In such cases we talk, not about changes *along* a demand (or supply) curve, but about *shifts* in the curve. For example, the invention of the electronic calculator had no effect on the cost of production of a slide rule or on the capacity of slide rule factories. Nevertheless, when the electronic calculator was invented the demand for slide rules dropped precipitously. We diagram that change by saying that the demand curve for slide rules shifted to the left. As Figure 2 suggests, if a shift to the left is dramatic enough, a product may simply cease to exist. If the lowest possible cost of producing a slide rule by the most efficient producer is $20, but even the consumer with the highest reservation price is unwilling to pay $20 (perhaps because she can obtain an equivalent electronic calculator for $16.00), then no one can make slide rules profitably. They will go the way of the quill pen, the vacuum tube, and the washboard. The shifted demand curve (D_2 in Figure 2) illustrates this: it never intersects the supply curve.

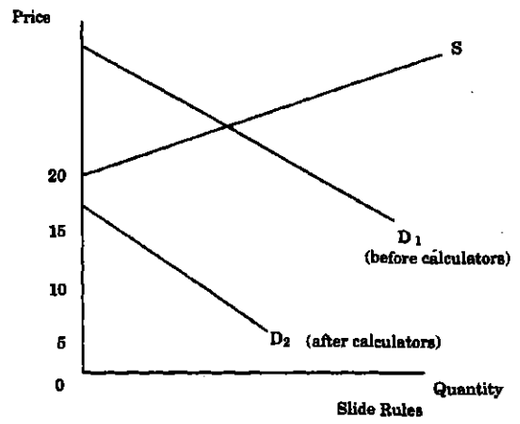

Figure 2

[609a]

6. See § 7.3.

7. Eastman Kodak Co. v. Image Technical Services, 504 U.S. 451, 112 S.Ct. 2072 (1992). See § 7.6a.

Supply curves may shift just as demand curves do. The invention of the microprocessor in a silicone chip reduced the cost of building computers by a factor of one hundred. The result is that the new supply curve for computers in the 1990's is much lower than the supply curve of the 1960's, and equilibrium output is much higher.

1.1b. Behavior of the Competitive Firm

We have considered the competitive, multi-firm market, and can now examine the behavior of the individual firm in that market. We assume a market with a large number of sellers, into which entry is relatively easy and can be accomplished in a short time. How will an individual firm in that market decide how much to produce and what price to charge?

Even though the steel market's equilibrium price is $3.00 per pound, there are still individual buyers, such as the orthodontists, whose reservation price is far higher than $3.00. Suppose that the individual firm attempts to charge a higher price than $3.00—perhaps $4.00—for a pound of steel. The orthodontists are certainly willing to pay $4.00, but if they can buy for $3.00 they will do so. When one firm in a 100–firm market attempts to charge $4.00, a buyer who knows that the "going" price is $3.00 will look for a different seller. In a perfectly competitive market in which all buyers have complete price information, all the sellers will be "price takers"—they must simply accept the market price as given. No single firm is large enough to influence either the total amount produced or the market price. As a result, the individual firm can sell as little or as much as it pleases at the market price, but it will lose all sales if it attempts to charge more.

The situation facing the perfect competitor can be described in two ways. First, the firm faces a perfectly horizontal demand curve, as is illustrated in Figure 3. For the perfect competitor the market price is the same at all rates of output. Alternatively, the individual competitor faces extremely high firm elasticities of supply and demand. In response to a very small price increase, alternative suppliers will immediately offer substitute products to the price raiser's customers, and all customers will switch to those substitutes. The firm will lose all of its sales.

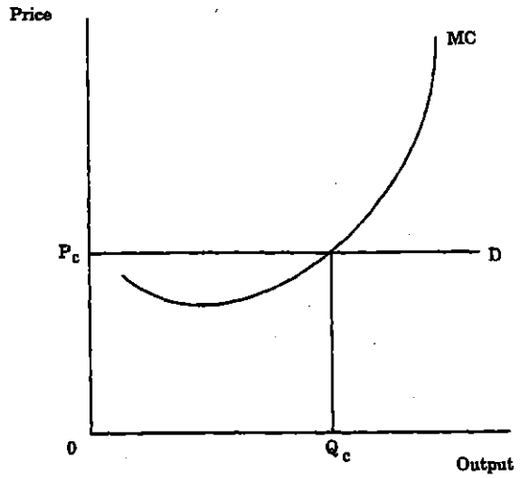

Figure 3

[610a]

One must therefore distinguish between *market* elasticities of supply and demand, and individual *firm* elasticities of supply and demand. Except for the pure monopolist (whose output is the same as the output of the entire market) the individual firm faces higher elasticities of supply and demand than does the market as a whole. This is because *within* a market substitution is easy and quick. If the market contains 100 producers of identical steel, then A's steel is indistinguishable from B's steel, which is indistinguishable from C's, and so on. The fact that customers are indifferent as to whose steel they buy means that they will switch immediately to B or C if A attempts to increase price; conversely, B or C will happily provide the steel.

The firms in a perfectly competitive market have little discretion about what price to charge. They do make individual decisions, however, about the amount to produce. Even in a perfectly competitive market with an established single market price, different firms are of different sizes and produce differing amounts.

The individual competitor's output decision is a function of its marginal costs. Marginal

cost is the additional cost that a firm incurs in the production of one additional unit of output. The best way to understand marginal cost is to consider several related cost curves. A firm's costs can be divided into two broad categories, fixed and variable. Fixed costs are those costs that do not change with output over the short-run, which is some finite period of time, usually less than the lifetime of the plant. Land costs, property taxes, management salaries, plant and durable equipment all generally fall into the category of fixed costs. Once the money for fixed cost items is invested it must be paid whether or not the plant produces anything, and the costs do not vary with the amount the plant produces.

Variable costs, by contrast, are costs that change with output. For the steel mill, the costs of iron ore and other raw materials are variable costs, as are fuel to burn in the refining furnaces, hourly wages, and transportation. If a firm increases its output by, say 10%, the cost of all these things rises because the firm must purchase more. The cost of the plant, durable equipment and the president's salary are likely to stay the same. Over the long-run, however, even these "fixed" costs must be considered variable. Eventually plant and durable equipment will have to be replaced. The firm will then decide whether to increase capacity, decrease it, or perhaps even go out of business.

Both fixed and variable costs are generally expressed as costs per unit of output. These are illustrated in Figure 4. "Average fixed cost" (AFC) is the amount of fixed cost divided by the amount of output. Since total fixed costs remain constant, average fixed costs decline as output increases. "Average variable cost" (AVC) is total variable cost divided by the amount of output the firm produces. The behavior of the average variable cost curve is more complex. Every established plant has some particular range of output in which it is most efficient. For example, a plant properly designed to produce 80–100 units per year will perform at lowest cost when output is in that range. If output drops to 50 the plant will perform less efficiently and per unit costs will rise. Thus the AVC curve shows higher than minimum AVC at low outputs. Blast furnaces, to give just one example, cost the same amount to heat whether they are used at capacity or only at half capacity. The same thing generally holds true for output that exceeds the plant's "optimal capacity." For example, a plant and work force designed to produce 80 units per week may be able to increase output to 100 units per week only if workers are paid overtime wages, which may be twice their normal wages, or if equipment is used at a level at which its breakdown rate is high. Thus, the AVC curve increases to the right of the minimum point as output increases.

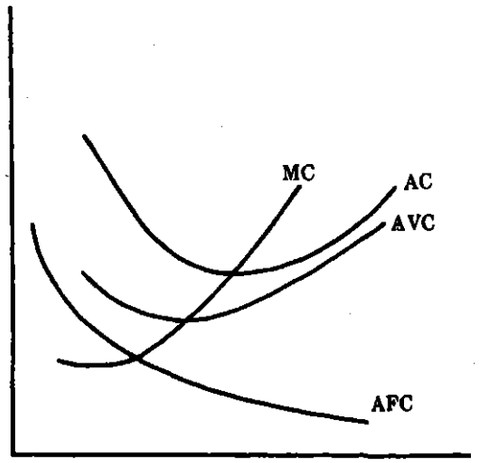

Figure 4

The average variable cost curve (AVC) of the plant tends to be U-shaped. Average variable cost declines as output increases toward the optimal output for the plant. AVC is at the lowest point when the plant is producing the optimal output for which it was designed, and increases when the plant's output exceeds optimal capacity. The AVC curve often has a relatively flat bottom, because many plants are efficient over a fairly broad range of output.

Just above the AVC curve in Figure 4 is the average total cost curve (AC), usually called the average cost curve, which is nothing more than the *sum* of all fixed and variable costs divided by output. Since all costs are either fixed or variable, the AC curve represents the total costs that a firm incurs. As a result the AC curve is important in determining the firm's profitability. In order to be profitable

the firm must obtain an average price per unit equal to or greater than AC. The AC curve is shaped roughly like the AVC curve, except that the two converge as output increases.[8]

Once again, *marginal* cost is the additional cost that a firm incurs in producing one additional unit of output.[9] Since a firm incurs no increased *fixed* costs in expanding output in the short run, marginal cost is a function of variable costs alone. The marginal cost curve (MC) falls and rises more dramatically than the average variable cost curve does, because the marginal cost curve considers merely the additional costs of one added unit of output. By contrast, the AVC curve averages that difference over the entire output being produced.[10] Importantly, the marginal cost curve always intersects the AVC curve at its lowest point. A minute's reflection about averages will tell you why. Suppose that you are averaging the height of United States Supreme Court Justices and you have managed to gather eight of them, and have computed their average height as 6'0". Now the ninth Justice walks in the door and happens to be 5'3" tall. The average height will decline. But if the ninth Justice happens to be 6'7" the average will increase. Whether the average falls or rises is a function of the height of the "marginal" Justice. As long as the marginal Justice is below the average, the average will decline; as soon as the marginal Justice is above the average, the average will increase.

The relationship between the cost curves is illustrated in Table A. Notice that as output increases total fixed costs remain constant at 120. As a result, average fixed costs decline steadily, but at a decreasing rate. Total variable costs increase constantly as output increases; however, they increase more slowly as the plant approaches optimal capacity (in the 3–5 output range), and more rapidly again as the plant exceeds optimal capacity. As a result, *average* variable cost bottoms out at an output of about 5 and then increases.

Output	Total Fixed Cost	Average Fixed Cost	Total Variable Cost	AVC	MC	AC
1	120	120	200	200	200	320
2	120	60	240	120	40	180
3	120	40	270	90	30	130
4	120	30	320	80	50	110
5	120	24	375	75	55	99
6	120	20	510	85	135	105
7	120	17.14	700	100	190	117.14

TABLE A

How will the competitive firm make its output decision? Suppose the market price is $100.00 per unit. At its current rate of production the firm has marginal costs of only $60.00 per unit. That is, if it produced one additional unit it would incur $60.00 in additional costs. The production of the additional unit will generate profits of $40.00. A profit-maximizing firm will increase production by one additional unit. However, suppose that the firm's marginal cost at its current rate of output is $120.00. If it produced one fewer unit it would spend $120.00 less. In that case the production of the last unit is generating $20.00 in losses: the firm could make $20.00 more by producing one unit less.

8. The AC and AVC curves converge because AC is equal to the vertical sum of AVC and AFC; as output increases AFC continually decreases, approaching zero.

9. Or, $MC = AC_j - AC_i$, where the difference between output i and output j at any level is one unit. In the short run, it is also true that $MC = AVC_j - AVC_i$; that is, short-run marginal cost is a function of variable costs alone.

10. For example, suppose that AVC for 100 units is 3, and at that point marginal cost is 6. When unit 101 is produced, marginal cost is 6, but AVC would rise only to 306/101, or 3.029.

Look back at Figure 3 to see the relationship between the competitive firm's marginal cost curve and the demand curve that it faces. The firm will always try to produce at a rate of output at which its marginal cost equals the market price. If it is producing more than that, it can increase profits by decreasing production. If it is producing less it can increase profits by increasing production. The competitive rate of output in Figure 3 is Q_c.

Two observations are important. First, although economists sometimes say that a firm's efficiency is a function of its marginal costs, all competitive firms have the same marginal cost at current output levels. If the current market price of widgets is $100.00, and the market is perfectly competitive, all firms at their current output rate will have marginal costs of $100.00. (If marginal costs never drop to $100.00, then the firm is so inefficient that it will not produce at all.) The efficiency differences show up, not in the marginal costs, but in the rate of output. That is, a more efficient firm will produce more units of output than a less efficient firm produces at the same marginal cost level.

Second, not every firm in a competitive market is necessarily profitable. The fact that every firm has a point on its marginal cost curve which is lower than $100.00 does not tell us anything about the firm's profitability when the market price is $100.00. In order for the firm to be profitable, that point on the marginal cost curve must be at or above the firm's average (total) cost curve. Even if the firm is losing money, however, if it produces at all it will produce at the rate at which price equals marginal cost. In that case that rate of output will be the "loss-minimizing" rather than the "profit-maximizing" rate of output.

Although the market price might be less than a firm's average total cost at any output level, the firm will not necessarily cease production. The fixed costs may have been "sunk"—that is, the firm may not be able to recover them if it goes out of business. Further, the fixed costs must be paid whether or not the plant produces. As a general rule, the firm will be able to cut its losses as long as the market price is above its average variable costs, and it will continue to produce. However, when the plant wears out and needs to be replaced, the firm may then decide to go out of business, or else to build a more efficient plant.

Perfectly competitive markets are generally thought to be "efficient" because they do the best job of providing consumers with goods at the cost of producing them. As a result, competition maximizes the total value of goods produced in society. In a competitive market no single firm has the power to reduce the available supply of goods, and no firm has the power to increase the price above the market level.

The world contains no perfectly competitive markets, and many markets do not even come close. Firms often differentiate their products from other firms; as a result, customers are no longer indifferent to the identity of the seller. Information about market conditions is always less than perfect; as a result many transactions take place at some price other than the market price, and some socially valuable transactions never occur at all.[11] "Economies of scale"—the ability of larger firms to produce at a lower cost than smaller firms—may result in markets that have fewer than the number of sellers required for perfect competition to occur.[12] In short, like all scientific models, the model of perfect competition applies only imperfectly in the real world; nevertheless it can be of great service to the antitrust policy maker in predicting the consequences of a certain action or legal rule.

§ 1.2 Monopoly

1.2a. Price and Output of the Protected Monopolist

The monopolist—the only firm selling in a particular market—faces a different array of

11. In general, the more expensive it is for consumers to search out relevant information about prices and markets, the more likely they will make a less than optimal transaction. As a result, prices tend to vary more in markets where search costs are high in relation to the value of the product. See G. Stigler, The Economics of Information, 69 J. Pol. Econ. 213 (1961); G. Stigler, The Theory of Price 2–6 (3d ed. 1966).

12. See § 1.4a.

price and output decisions than those that confront the perfect competitor. For this formal analysis we assume that the market contains only one firm, whose demand curve is therefore identical with the market demand curve. Second, the formal monopolist does not need to worry about new entry by a competitor. These assumptions often will not apply to the *de facto* "monopolist" that exists in most antitrust litigation. The antitrust "monopolist" is a dominant firm, but the market may contain a competitive "fringe" of smaller competitors.[1] Second, the antitrust monopolist ordinarily has no legal protection from competitive entry. If either formal assumption is relaxed the monopolist will face a certain amount of "competition" and will vary its behavior accordingly.[2] Assuming, however, that the monopolist has a 100% share of a market and no concern about entry by a competitor, how much will it sell and what price will it charge?

The monopolist has one power that the perfect competitor does not have. If the monopolist reduces output, total market output will decline, for the monopolist is the only producer in the market. As total market output goes down, the market-clearing price goes up. As a result, the monopolist, unlike the competitor, can obtain a higher price per unit of output by producing less.

However, the monopolist will not be able to charge an infinite price for its product. Even the orthodontists may be unwilling to pay more than $3000 per pound for steel; if the price goes higher they will change to silver or some other alternative.

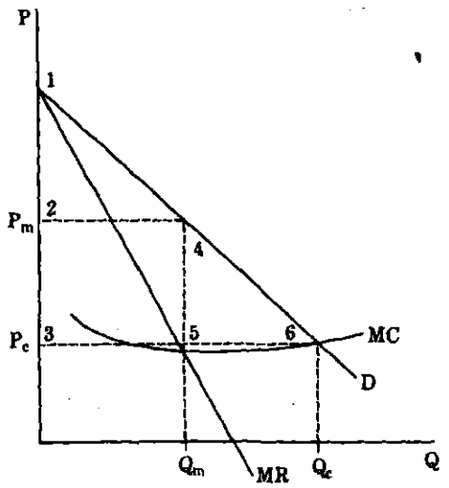

Figure 5

Figure 5 introduces the marginal revenue curve (MR), which represents the additional revenue that the monopolist obtains when it produces one additional unit of output. As Figure 5 shows, the marginal revenue curve facing the monopolist is steeper than the demand curve.[3] This is because the monopolist must sell all units of output at the same price. Thus the marginal revenue curve shows not only that increases in market output reduce the market clearing price (which is what the demand curve shows), but also that less revenue is obtained from sales of all units, not just the incremental unit. This is easy to see in Table B. At output of one unit, the price is $20 and the seller's marginal revenue—the difference between the amount it obtains from one unit and the amount it obtains from zero units—is also $20. When output increases to 2 units, price drops to $18. However, the monopolist must sell both the first and the second units for $18.00. While the price drops by $2.00, marginal revenue drops by $4.00—$2.00 for each of the two units. This process continues and yields the MR curve in Figure 5.[4]

§ 1.2

1. Economists generally speak of such firms not as "monopolists," but rather as "dominant firms."

2. For example, it may charge a lower, or "limit" price, calculated to make entry by outsiders less attractive. See § 8.3b.

3. If the demand curve is linear, the marginal revenue curve is also linear and exactly twice as steep as the demand curve. For a simple proof, see F. M. Scherer & D. Ross, Industrial Market Structure and Economic Performance 21 & n. 13 (3d ed. 1990).

4. Marginal revenue can also be expressed as:

$MR = \delta R/\delta Q$,

where δR equals the change in total revenue and δQ equals the corresponding change in quantity. As a result, marginal revenue can also be written as $MR = P + Q(\delta P/\delta Q)$, where P equals the price paid by the marginal consumer, and $\delta P/\delta Q$ equals the change in price necessary to attract the marginal consumer. For example, assume that at a price of 11, 20 units are sold; at a price of 10, 21 units are sold. In that case marginal revenue per unit, going from a price of 11 to a price of 10 equals:

$10 + 20(-1/1) = -10$

That is to say, marginal revenue equals the price paid by the new, or marginal consumer (10), plus the change in revenue that accrues as a result of the price change to the 20 existing consumers. In this case, the twenty existing customers pay $1 less apiece.

The profit-maximizing monopolist, just as the profit-maximizing competitor, will expand production to the point that one additional unit will produce greater additional costs than additional revenues. It will produce at point Q_m on the graph in Figure 5 and charge price P_m. If the monopolist expands output beyond Q_m the additional revenue, shown by the MR curve, will be less than the additional costs, shown by the marginal cost curve (MC). P_m is known as the "monopoly price," or as the monopolist's "profit-maximizing price."

Output	Price	Total Revenue	Marginal Revenue
1	$20	$20	$20
2	$18	$36	$16
3	$16	$48	$12
4	$14	$56	$8
5	$12	$60	$4
6	$10	$60	0
7	$8	$56	$-4
8	$6	$48	$-8

TABLE B

Both the perfect competitor and the monopolist maximize profits by equating marginal revenue and marginal cost. For the competitor, the marginal revenue curve is identical with the demand curve, and therefore with the market price.[5] For the monopolist, by contrast, the marginal revenue curve and marginal cost curve intersect to the left of the marginal cost curve's intersection with the demand curve. The monopolist produces at a lower rate than would a perfect competitor in the same market, and its profit-maximizing price is higher.

The difference between the monopolist's profit-maximizing price, P_m, and the competitor's profit-maximizing price, P_c tells us something about the degree of power that the monopolist has. If P_c is $1.00, a monopolist whose profit-maximizing price is $1.50 has more monopoly power than one whose profit-maximizing price is $1.02. The Lerner Index, discussed in § 3.1a, expresses market power in this way through the use of a simply derived formula relating the firm's marginal cost to its profit-maximizing price.

A monopolist's market power is a function of the elasticity of demand for its product. If the elasticity of demand for pistachios at the competitive price is high, consumers will be sensitive to changes in the price. If the price goes too high many will buy a substitute, such as almonds or cashews. In that case, the "spread" between the competitive price and the monopolist's profit-maximizing price will be relatively small. However, if the elasticity of demand is low, then consumers view the product as having few good substitutes. The monopolist will be able to extract a much higher price without losing too many sales.

Market power can also be computed directly from a firm's price elasticity of demand. The formulas are also discussed in § 3.1. The formulas offer several insights about the relevant variables in market power measurement. In general, however, the formal analysis of market power is of little use to a court because the elasticity of demand a firm faces cannot be computed in litigation.

1.2b. Monopsony; Output Effects; Policy Implications

The mirror image of monopoly is "monopsony." A monopsonist is a monopoly buyer

This equation enables us to relate a firm's market power to the elasticity of demand facing it. See the discussion of technical measurement of market power in § 3.1a.

5. This is so because the competitive price remains constant at all rates of output. For example, if price is $20, each additional sale at any output level the competitive firm chooses will generate an additional $20, and marginal revenue will remain constant at $20.

rather than seller. Although most antitrust litigation of market power offenses has involved monopoly sellers rather than buyers, monopsony can impose social costs on society similar to those caused by monopoly.[6]

By reducing its demand for a product, a monopsonist can force suppliers to sell to it at a lower price than would prevail in a competitive market. Some people are skeptical about this conclusion. No supplier would stay in business if it were forced to sell to the monopsonist at a price lower than its average costs, and price would tend toward average cost in a competitive market. Can a monopsonist actually force suppliers to engage in continuous loss selling?

The answer is no, of course. However, not all suppliers have the same costs, and many suppliers will have lower average costs if they reduce their output. When the price in a competitive market is $1.00, then the average costs of the least efficient, or "marginal," supplier are near $1.00. However, there may be other sellers who have lower costs. If the monopsonist announces that it will pay only 90¢ in the future, then the marginal sellers in the market—those with costs in the 90¢ to $1.00 range—will drop out, at least if the 90¢ price persists and they are unable to reduce their costs. Likewise, when prices are at the competitive level most firms have a rising marginal cost curve. If the price is suppressed they will reduce output to a level that once again equals their marginal costs. In any event, both price and output will fall below the competitive level when the buyer is a monopsonist. Some productive assets will be assigned to products that would have been the supplier's second choice in a competitive market. As a result, monopsony allocates resources inefficiently just as monopoly does.[7]

The important policy implication of monopsony is that it *reduces* rather than increases output in the monopsonized market. Many federal judges have failed to see this. The consumer welfare principle in antitrust, or the notion that the central goal of antitrust policy should be low prices,[8] has often suggested to courts that monopsony is not all that important an antitrust policy concern. For example, in *Balmoral* the court faced an agreement among theater operators not to bid against each other for motion pictures.[9] As a result, the prices they paid for the pictures were lower than if they had bid competitively. The court suggested that such an agreement could result in lower prices to consumers and concluded that the agreement might "serve rather than undermine consumer welfare."[10] Likewise, in the *Kartell* case,[11] the First Circuit refused to condemn as monopolistic a health insurer's policy of setting the maximum price it was willing to pay for health care services used by its insureds. The court noted that "the prices at issue here are low prices, not high prices * * *. [T]he Congress that enacted the Sherman Act saw it as a way of protecting consumers against prices that were too *high*, not too low."[12]

These decisions suggest that monopsony buyers will generally pass their lower costs on to their consumers. But that is not necessarily the case. The monopsonist reduces its buying price by *reducing* the amount of some input that it purchases. If the input is used in the output in fixed proportions, then the output must be reduced is well. This suggests two things: (1) the monopsony buyer that resells in a competitive market will charge the same price, but its output will be lower than if it were a competitive purchaser; (2) the monopsony buyer (or cartel) that resells in a monopolized (or cartelized) market will actually charge a *higher* price than if it were a competitive purchaser.

Consider this illustration. A monopoly manufacturer of aluminum is also a monopsony purchaser of bauxite. Bauxite is an ingredient in aluminum, and one ton of bauxite, when mixed with other ingredients, yields two tons of aluminum. In a competitive market bauxite sells for $25 per ton and the producer would purchase 1000 tons, which it would then use to make 2000 tons of aluminum. The aluminum would be sold at the monopoly price of $80 per ton. In the monopsonized bauxite market, however, the monopsonist/monopolist reduces its purchases of bauxite to 700 tons, which it purchases at $20 per ton. If it uses bauxite and other ingredients in fixed proportions of one ton of bauxite to two tons of aluminum, then it must also reduce the output of aluminum to 1400 tons. In that case, the market clearing price of the aluminum will rise to, say, $105.00. In sum, even though the monopsonist/monopolist buys an input at a lower price, the lower output entails a higher, not a lower, resale price.[13] If the monopsonist/monopolist can change the proportion of bauxite in its aluminum the story becomes more complicated. But in general two things will be true. First, the price of aluminum will not go down and will almost always go up anyway. Second, consumers will not get the aluminum alloy that they would have gotten in a competitive market.[14]

6. For a thorough, readable study of the law and economics of monopsony, see R. Blair & J. Harrison, Monopsony: Antitrust Law & Economics (1993). On the law of buying cartels, see 12 Antitrust Law ¶ ¶ 2010–2015 (1999).

7. There is some ambiguous legislative history suggesting that Senator Sherman did not intend his proposed statute to apply to monopsony or buyers' cartels.

> Senator George (D.Miss.): Upon the formation of [the] bagging trust the cotton farmers * * * agreed that they would not purchase jute bagging, and by that agreement * * * the rich rewards anticipated by the * * * trust were defeated. The fact that the bill * * * applied to all arrangements * * * by whomsoever made, would bring within its reach all defensive agreements made by farmers for the purpose of enhancing the price of their products * * *.
>
> * * *
>
> Senator Sherman: That is a very extraordinary proposition. There is nothing in the bill to prevent a refusal by anybody to buy something. All that it says is that the people producing or selling a particular article shall not make combinations to advance the price of the necessaries of life.

20 Cong.Rec. 1458 (1889).

8. See §§ 2.2–2.3.

9. Balmoral Cinema v. Allied Artists Pictures Corp., 885 F.2d 313 (6th Cir.1989).

10. Id. at 317. The court approved the lower court's instruction to the jury to apply the rule of reason, and its subsequent judgment for the defendants.

11. Kartell v. Blue Shield (Mass.), 749 F.2d 922 (1st Cir.1984), cert. denied, 471 U.S. 1029, 105 S.Ct. 2040 (1985).

12. Id. at 930–931.

13. Although the monopsonist purchases at a lower absolute price, it has a higher effective marginal cost (actually, marginal outlay) than the buyer in a competitive market. Each incremental unit that the monopsonist purchases, assuming it cannot price discriminate in its buying, entails a higher price for all previously purchased units as well. For example, assume that if the monopsonist purchases 100 units the price is 25¢, but if it purchases 101 units the price rises to 26¢. The marginal outlay for the move from 100 units to 101 units is 100 × 1¢ + 26¢, or a total of $1.26. By contrast, the marginal cost of unit 101 is only 26¢. "Marginal outlay" refers to the total additional cost that the monopsonist incurs when it purchases one more unit. By contrast, "marginal cost" refers to the cost of the one additionally purchased unit. While the monopolist generally maximizes profits by equating marginal cost and marginal revenue, the monopolist that is also a monopsonist in an input market maximizes profits by equating marginal outlay and marginal revenue.

Figure 6 illustrates. It shows the relevant demand (D), marginal revenue (MR), marginal cost (MC) and marginal outlay (MO) curves of a firm that purchases a single input in a monopsonized market and resells this input in a monopolized market. Considering the firm simply as a monopolist in the output market, it would equate MC and MR. The monopoly price would be P_m and monopoly output would be Q_m. However, if the monopolist is also a monopsonist in the market for the input and its marginal cost curve slopes upward, then its marginal outlay curve will slope upward as well, only twice as steeply. That is, the relation between marginal cost and marginal outlay is exactly the same as the relation between demand and marginal revenue, except turned upside down. The monopolist/monopsonist maximizes its profits by equating MO and MR. This yields a monopoly/monopsony price of P_{mm}, and an output of Q_{mm}.

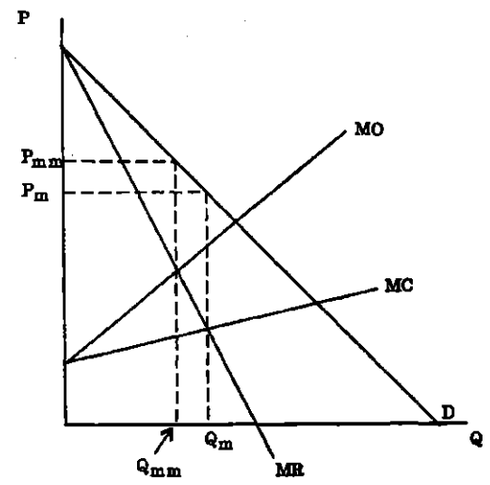

Figure 6

[611a]

14. For a more technical explanation, see R. Blair & J. Harrison, Antitrust Policy and Monopsony, 76 Cornell L. Rev. 297, 299–300 (1991).

The foregoing suggests two things. First, monopsony is an important antitrust concern and is just as inconsistent with consumer welfare as monopoly is. Indeed, one should *never* presume that the lower prices paid by a monopsonist are passed on to consumers as lower resale prices. Second, however, the antitrust policy maker must distinguish between lower buying prices that result from reduced transaction costs or the elimination of upstream market power, and lower buying prices that result from monopsony. If a large buyer is able to obtain lower prices by reducing transaction costs, the buyer will generally buy *more* rather than less.[15] The result will be lower resale prices, even if the large buyer resells in a monopolized market. Further, as § 9.2 illustrates, the firm that purchases at a lower price by eliminating an upstream monopolist or cartel virtually always charges a lower price on resale. Once again, this is true for both the competitor and the monopolist in the resale market.

A principal difficulty of antitrust policy toward monopsony is distinguishing between the efficient low purchase prices that result from reduced transaction costs or elimination of upstream monopoly, and the inefficient low purchase prices that result from monopsony. Perhaps the most problematic area is joint purchasing arrangements, which create a significant potential for cost savings but may also facilitate buyer price fixing.[16] In such a case the decision maker should try to determine whether the defendants' managers are encouraging members to purchase as much as possible, which is generally inconsistent with buyer price-fixing; or encouraging them to suppress their buying, which is highly suspicious.

1.2c. De Facto Monopolies in Real World Markets

The analysis of monopoly in this section was predicated on two assumptions—namely, that the monopolist had 100% of its market and that new entry was impossible. Such monopolies do exist in the real world, but most of them are price-regulated public utilities, such as electric companies. The rationale for the legal recognition of such "natural monopolies" is discussed below in § 1.4. Most antitrust policy concerning monopolies is directed at the *de facto* monopolist, which has no such legal protection. The *de facto* monopolist most generally does not have 100% of its relevant market, although the percentage may be close. Furthermore, the *de facto* monopolist must consider the possibility of entry by new firms.

Once these two assumptions of pure monopoly are relaxed, analyzing the monopolist's output and price decisions becomes more difficult. The *de facto* monopolist behaves strategically. In making a price or output decision it must either take the current output of competitors into account, or else it must try to anticipate responses by small competitors or potential competitors. It may also strategize a price or output decision designed to eliminate a competitor or potential competitor from the market. Much of antitrust law is concerned with the strategic decisions of the *de facto* monopolist trying to enlarge or protect its monopoly position.

The *de facto* monopolist may deter or delay competitive entry by setting a lower price than the one determined by the intersection of its marginal cost and marginal revenue curves. In general, the *de facto* monopolist has two choices. On the one hand, it can forget about new entry and earn as much as possible right now. In that case the monopolist will make maximum monopoly profits in the immediate future, but its monopoly position will be more quickly eroded by competitors and new entrants who are attracted by the high profits. On the other hand, the monopolist might set a lower "entry deterring" or "limit" price. Under limit pricing the monopolist will earn lower profits today, but its stream of profits will last longer, because new firms will not be as eager to enter the market. Which alternative the monopolist takes varies from case to case, and economists have different opinions about the circumstances under which each will occur.[17] Much of this debate is relevant to antitrust policy, and is discussed in chapters 6–8.

Whichever decision the firm makes, it will likely be attempting to maximize its profits. That is to say, a firm's monopoly profits are a function not only of their magnitude at any instant, but also of their duration. Thus we speak of a "short run" profit-maximizing price, determined by the intersection of marginal cost and marginal revenue, which maximizes the monopolist's profits in the immediate instant. But we can also speak of a "long run" profit-maximizing price that takes the duration of monopoly profits into account as well. The latter price will often be significantly lower than the former.

§ 1.3 Antitrust Policy and the Social Cost of Monopoly

1.3a. Monopoly as a Status; Monopolization as a Process

A *social cost* is a net loss that society suffers as a result of a particular transaction. A *social benefit* is a net gain. If A gives B $100, B is $100 richer and A is $100 poorer. Disregarding the costs of the transaction itself, such "transfer payments" produce neither a social cost nor a social benefit. By contrast, if A produces for $100 a widget that B values at $150, society may become $50 richer. B might pay $150 for the widget. In that case B will be neither better nor worse off, for he valued the widget by just what he paid for it. But A will be $50 richer, for his costs were only $100. Alternatively, if A sells the widget at $100, A will be neither better nor worse off, but B will be $50 better off.

If A holds out for a price of $150 and B is willing to pay only $140, however, the transaction will not occur. In that case no one will be better off. B may then enter into a transaction with C and purchase a substitute that B values at perhaps $130, and which costs C, say, $110. The price will be between $110 and $130. Even if that alternative transaction occurs, however, society will be only $20 better off. The substitute transaction is less favorable to both B and society as a whole than B's preferred transaction would have been.

Social costs can also result when transactions injure someone who was not a party to the transaction. For example, the builder of a factory may not bother to negotiate with neighbors for the right to pollute their air, particularly if he thinks the neighbors have no legal right to protect their air from pollution. However, the neighbors are worse off. The common law of nuisance and the National Environmental Policy Act are both attempts to force the factory to "internalize" and pay at least a part of this cost.[1]

For antitrust purposes, the *social cost* of monopoly is equal to the loss produced by monopoly pricing and monopoly behavior, minus any social gains that monopoly produces. *Monopolization*—or the antitrust offense of creating or maintaining a monopoly by means of anticompetitive exclusionary practices—is a process rather than merely an outcome. We sometimes distinguish the two when we call

15. See, for example, Northwest Wholesale Stationers v. Pacific Stationery and Printing Co., 472 U.S. 284, 105 S.Ct. 2613 (1985), where the Supreme Court noted that joint buying is most generally efficient—a claim that the courts can test by assessing the venturers' market share of the market in which they buy. See also All Care Nursing Service, v. High Tech Staffing Services, 135 F.3d 740 (11th Cir.1998), cert. denied, ___ U.S. ___, 119 S.Ct. 1250 (1999), which approved a joint venture among hospitals to take bids from nursing service providers. The court cited the absence of any evidence of market power.

16. See § 4.1d. See also 13 *Antitrust Law* ¶ 2135 (1999).

17. For example, if entry will occur in ten years at any price higher than marginal cost, the incumbent would be best off to charge its short-run profit-maximizing price. So a patent monopolist, for example, who knows entry will occur immediately after its patent expires would probably equate marginal cost and marginal revenue today. By contrast, a firm whose short run profit-maximizing price of $1.50 will encourage entry in one year, while a price of $1.40 will delay entry indefinitely, will likely charge the latter price. The less a firm knows about the rate at which others will enter, or the more volatile the market, the more likely that the firm will charge its short-run profit maximizing price. In such cases the value of entry-deterrent pricing must be discounted by the uncertainty of the profits it will produce over the future.

On limit pricing, see W. Kip Viscusi, John M. Vernon & Joseph E. Harrington, Jr., Economics of Regulation and Antitrust 166–179 (2d ed. 1995).

§ 1.3

1. See 42 U.S.C.A. §§ 4321–47; and see R. Coase, The Problem of Social Cost, 3 J.L. & Econ. 1 (1960); A.M. Polinsky, An Introduction to Law and Economics 11–26 (2d ed. 1989).

the outcome "monopoly," and the process by which it is created by a term such as "monopolization," or "rent seeking." For any antitrust policy concerned with minimizing the social cost of harmful activity, both the process and the outcome are properly counted as a part of the activity's social cost, and part of the reasons for prevention. This is generally true of the economic theory of criminal behavior. For example, the social cost of theft is not merely the money value of the stolen object—indeed, the theft itself is only a wealth transfer. The social cost must also include the collateral damage that the thief inflicts on society, as well as the costs of the elaborate mechanisms that we use to deter theft.[2]

To be sure, some of the processes that create monopoly are efficient. For example, monopoly can be created by research and development. So we must have rules that distinguish harmful from beneficial practices that create monopoly. But this problem of definition or characterization is quite different from the question whether losses caused by harmful exclusionary practices should be counted as part of monopoly's social costs.

The *policy* question of monopoly's social cost always trades off relative gains and losses. Every state of affairs includes some positive social costs. Even vigorous competition entails costs that monopoly might avoid, such as the costs of making and interpreting competitive bids, or the inefficient duplication of productive assets or processes. One can always imagine a system with lower social costs than the present situation. So when we ask whether something is a social cost, we must always consider "relative to what?"

The earliest measures of the social cost of monopoly in the American economy took a kind of "public utility" approach to monopoly. The authors dealt with monopoly as if it were a given equilibrium condition, giving no consideration to the method by which the monopoly was created or preserved, or the mechanisms by which it might eventually be destroyed. In such a static situation, the only social cost of monopoly is the "deadweight" loss that it produces—a loss caused principally by the fact that consumers make inefficient substitutions in order to avoid paying monopoly prices.

But antitrust law is not frequently concerned with such equilibrium monopolies, for they are generally the product of legislation. Further, even those that result from legislation can impose social costs that the traditional deadweight loss triangle fails to capture. For example, if the owner of a shopping mall bribes a city council into refusing to rezone nearby property at the request of a potential competitor, the social cost of the monopoly will be (1) the deadweight loss caused by the incumbent's monopoly output restriction and price increase; (2) at least part of the expenses paid by the shopping mall owner in influencing the city council, and (3) the investment in planning a competing development that the potential entrant will now lose as a result of the incumbent's bribery.

Antitrust law's concern with this *process* of monopolization, rather than merely with the outcome, is quite apparent from the statutory scheme. The law of monopolization requires not only a monopoly position, but also the commission of one or more anticompetitive "exclusionary practices," thus signalling that the process by which monopoly is to be created determines its legality.[3] We condemn collusion, attempts and conspiracies to monopolize, tying arrangements, exclusive dealing, mergers and other practices only because we believe that these tend to facilitate the creation of monopoly. We may sometimes be wrong about our underlying facts or even about the economic theories we employ, but the basic premise remains the same: the principal target of the antitrust laws is not static monopoly as such, but rather the manifold mechanisms by which monopoly is created or preserved. Indeed, there is no law of "no fault" monopoly; the innocent monopolist does not violate the antitrust laws simply by charging its profit-maximizing price.[4]

2. See G. Becker, Crime and Punishment: An Economic Approach, 76 J.Pol.Econ. 169 (1968).

3. See chs. 6–8.

One possible explanation of antitrust's focus on process is that the real concern of the antitrust laws is the final outcome, but we need to deter, and deterrence is most effective if we hit things in the process of their creation. But the very fact that we fail to condemn the completed result *ipso facto* belies this claim. Antitrusters often say that their principal concern is monopoly, but that is not quite true. Their principal concern is monopoly created by certain means. Indeed the costs of the means by which monopoly is created and preserved may dwarf the costs of any misallocation caused by the monopoly pricing and output reduction themselves.

With these premises in mind, let us consider first the social cost of monopoly as economists have defined it, and then the expanded concerns of American antitrust policy with the social cost of monopoly's consequences and of the means by which monopoly is created and maintained.[5]

1.3b. The Deadweight Loss Caused by Monopoly

Monopoly forces some people to forego the transaction that was their first choice and would have produced the largest benefit. Rather, they take their second choice, which produces a smaller benefit.

Although monopoly imposes a social cost, society is not necessarily poorer because the monopolist exists. For example, society was clearly better off because Alcoa existed than if no aluminum producer existed at all, even if for many years Alcoa was a monopolist. Monopolist Alcoa produced a product that buyers valued more than the cost of producing it. Otherwise there would have been no market for aluminum. We talk about the "social cost" of the aluminum monopoly in order to underscore the fact that the production and sale of aluminum would have produced even greater social benefits had the market been competitive. The social cost of monopoly is the difference in social value between a monopolized market and a competitive market. It is not the difference in social value between a monopolized market and no market at all. For that reason the patent laws may be socially valuable, even though they create monopolies.[6]

In Figure 7, on page 20, P_c and Q_c show price and output in a competitive market. P_m shows the price for the same product in a market dominated by a monopolist, and Q_m shows the monopoly rate of output. Rectangle 2–3–5–4 represents a wealth transfer to the monopolist (the monopolist's output multiplied by the difference between the monopoly and competitive prices). Triangle 1–2–4 at the top of the diagram represents consumers' surplus, which is substantially less than it would be in a competitive market, where it would be triangle 1–3–6.

4. See 3 Antitrust Law ¶ ¶ 630–650 (rev. ed. 1996); and see § 6.3.

5. See H. Hovenkamp, Antitrust's Protected Classes, 88 Mich.L.Rev. 1 (1989).

6. See L. Kaplow, The Patent–Antitrust Intersection: A Reappraisal, 97 Harv.L.Rev. 1813 (1984).

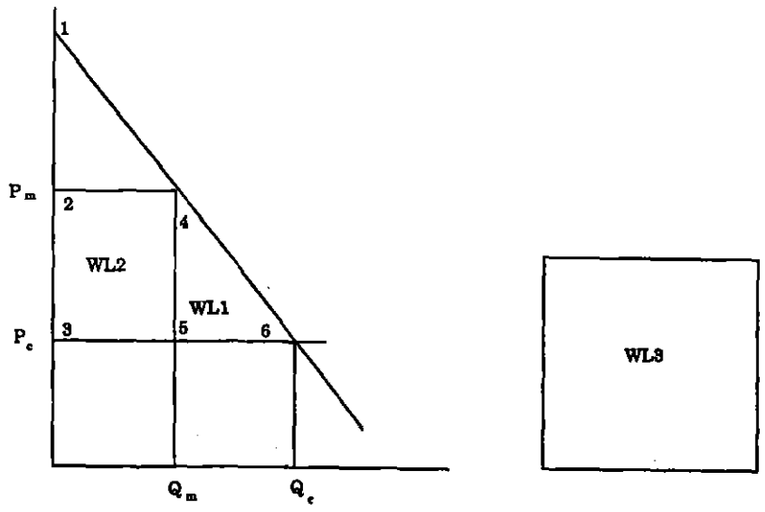

Figure 7

(620a)

Finally, triangle 4–5–6 represents the "deadweight loss" of monopoly. Consumers located along the demand curve between points 4 and 6 are not willing to purchase the monopolized product at the monopoly price, even though they are willing to buy it at the competitive price. Instead, they substitute to something that would have been their second choice in a competitive market. This inefficient substitution is traditionally spoken of as the social cost of monopoly. It is labeled "WL1," or welfare loss 1, in the Figure, for it is the oldest and most universally recognized of monopoly's social costs.[7]

Importantly, the traditional deadweight loss of monopoly does not derive from the fact that consumers pay higher prices. Within the pure monopoly model that loss to consumers is offset by an equal gain to the monopolist and from an efficiency standpoint is a "neutral" transfer of wealth. The deadweight loss arises because the monopoly encourages some customers to engage in an alternative transaction that produces less social value than would their first choice. A monopoly in the brick market may force a builder to switch to aluminum siding, even though he preferred bricks and was willing to pay the competitive price for them.

1.3c. The Social Cost of Monopoly: Rent–Seeking

At one time economists regarded triangle 4–5–6 as the only social cost of monopoly. But triangle 4–5–6 may understate the social cost of monopoly in real world markets. The discussion of monopoly in the previous section assumed that the monopolist was unconcerned about competitive entry. When that assumption is removed, as it is for *de facto* monopolists, then the social costs of monopoly are likely to loom larger.[8]

The *de facto* monopolist—the firm that does not have legal protection from new entry—must continually exclude competitors, who would increase output and drive prices down to the competitive level. In fact, the more profitable the monopoly, the more that potential entrants will be willing to spend in

7. A large literature on the size of the deadweight loss triangle of an individual monopolist and on the total deadweight loss caused by monopoly in the American economy is summarized in F. Scherer & D. Ross, Industrial Market Structure and Economic Performance 661–665 (3d ed. 1990).

8. See R. Posner, The Social Costs of Monopoly and Regulation, 83 J.Pol.Econ. 807 (1975). Other literature is summarized in Hovenkamp, Antitrust's Protected Classes, note 5.

order to enter the market, and the more the monopolist will spend to keep them out. Part or perhaps even all of rectangle 2–3–5–4 in Figure 7, which we characterized as a "wealth transfer," may not be a wealth transfer at all because the monopolist uses it up in entrenching its monopoly. At the outer limit the monopolist would spend *all* its expected monopoly profits in protecting its position, and would end up with no more than a competitive rate of return. This rectangle is labeled "WL2" in Figure 7.

With a linear demand curve and constant marginal costs, the area of the WL2 rectangle is precisely double the area of WL1.[9] But this hardly suggests that WL2 losses are always twice as large as WL1 losses. First, if the marginal cost curve is not horizontal (usually, it is rising through this range) and if the demand curve is nonlinear, then WL2 could be either less than or greater than WL1. Second, a monopolist does not necessarily spend all of WL2 in inefficient exclusionary practices. Presumably, at least part is paid to the owners as monopoly profits; another part is paid in efficient, rather than inefficient, attempts to secure or perpetuate the monopoly. In order to quantify the true social cost of monopoly, we must know something about how the monopolist spends these resources. So WL2 is best characterized as the outer limit of welfare losses of this type.

As the previous section observed, one way the monopolist might deter competition is by charging a price lower than its short-run profit-maximizing price. Although full analysis of such entry-deterring pricing is complex,[10] the short-run consequence is to make both the "wealth transfer" rectangle and the "deadweight loss" triangle smaller than they would be under short-run profit-maximizing pricing. Whether such entry-deterring pricing reduces the social cost of monopoly in the long run, however, depends on the effect of the pricing on the duration of the monopoly. A large deadweight loss that lasts for one year may still be less costly than a relatively small deadweight loss that lasts for ten.

A firm might also deter new entry by spending part of its monopoly profits in research and development (R & D), thus keeping ahead of its industry and making it more difficult for competitors to keep up. Throughout the 1970's, for example, IBM Corp. probably retained a dominant position in the computer market by being an aggressive innovator.[11] R & D may reduce the net deadweight loss of monopoly if society values the product of the R & D by an amount that exceeds its costs plus the increased social costs of any additional monopoly power that the R & D creates. Nevertheless, one effect of R & D will be to make new entry by competitors more difficult.

The relationship between R & D expenditures and monopoly is controversial, and has produced a number of conflicting theories. At one end is Joseph Schumpeter's argument that since research is both expensive and risky, firms in competition will not be able to afford it. A large amount of money spent without an assured return may be enough to deter a competitive firm from innovating.[12] A common rebuttal to this argument is that all new investment entails risk. Some research investments are rational and others are not. The consequences for a competitive firm of falling behind other competitors are just as serious as the consequences of spending R & D money unprofitably. Furthermore, competitors research in order to *acquire* market power. If they can somehow distinguish their product and make it more attractive than the product offered by others, the difference may show up as monopoly profits.

The monopolist unconcerned about competitive entry may not innovate very much. For

9. This is true because when the demand curve is linear, the marginal revenue curve is linear and twice as steep as the demand curve. If MC is also linear, this generates a "deadweight loss" triangle and a "wealth transfer" rectangle that have precisely the same base and height, but the area of a right triangle is one-half of base times height.

10. See § 8.3b.

11. But see L. Sullivan, Monopolization: Corporate Strategy, the IBM Cases, and the Transformation of the Law, 60 Tex.L.Rev. 587 (1982).

12. See J. Schumpeter, Capitalism, Socialism, and Democracy 106 (3d ed. 1950).

example, a monopoly public utility may have little incentive to innovate, particularly if cost-saving technology will reduce the base from which its rate of return is calculated.[13] By contrast, a monopolist threatened by competitive entry may spend a great deal on entry-deterring innovation. It has been argued that the monopolist may even engage in inefficient "predatory" innovation—that is, innovation reasonably calculated to preserve the monopolist's dominance, and whose monopoly efficiency losses will exceed any efficiency gains that result from the innovation itself.[14] Whether or not this theory has any economic merit, it has been popular among antitrust plaintiffs. Many monopolization cases in the late 1970's and early 1980's involved allegations that the defendant injured the plaintiff or drove it out of business by predatory product innovation.[15] The claim is still recognized today, although less frequently.[16]

Arguments have also been made that large firms can engage in research more cheaply than small firms because the larger firm can distribute the costs of R & D over a larger volume of production.[17] Likewise, a firm that operates in many markets might profit more from research and development than a single market firm because often research yields unanticipated or tangential benefits in markets other than the one for which it was undertaken. Neither of these arguments, it should be noted, depends on the firm's market power, but only on its large absolute size or on its operation in many markets. Furthermore, both arguments tend to be undermined by the fact that literally thousands of small firms engage daily and profitably in relatively sophisticated types of research and development. The computer revolution of the late 1970's and early 1980's, for example, involved the research activity of many tiny firms. Finally, there is a healthy market for the products of innovation. A small firm that is unable to take advantage of the consequences of innovation in an adjacent market will probably be able to license the innovation to someone else who can.

The ambiguous relationship between monopoly and innovation has been apparent in the case law since soon after the Sherman Act was passed. In the *American Can* case,[18] the court faced the defense that a monopoly created by merger should be preserved because the monopolist could afford research and development activities that had not occurred before the monopoly came into existence. The judge was "reluctant to destroy so finely adjusted an industrial machine * * *." Thirty years later Judge Learned Hand wrote that monopoly was bad because it "deadens initiative * * * and depresses energy," and because "immunity from competition is a narcotic, and rivalry is a stimulant, to industrial progress." In the very same opinion, however, Judge Hand found that Alcoa had illegally monopolized the market because it aggressively "embrace[d] each new opportunity as it opened" and faced "every newcomer with new capacity already geared into a great organization, having the advantage of experience, trade connections and the elite of personnel."[19]

Before criticizing judges for being unclear about the relationship between monopolization and innovation, however, one should note that economists have not done much better. Even today there is widespread disagreement about whether monopoly encourages or discourages research and development and, if monopoly encourages development, whether that fact increases or decreases the social costs of monopoly.[20] No easy generalizations have been forthcoming.

The monopolist threatened with new entry may also spend part of its monopoly returns in less ambiguous entry-deterring practices which increase the social costs of monopoly. Properly defined predatory pricing,[21] sabotage, espionage, vexatious litigation,[22] false and misleading advertising can all have the effect of prolonging the period during which a *de facto* monopoly exists and thereby increase the social cost of the monopoly.

Monopoly may also yield certain inefficiencies that are not planned but which appear to accompany the absence of competition in a market. For one thing, the monopolist is a "price maker" rather than a "price taker." The monopolist, unlike the competitor, must calculate its profit-maximizing price by predicting how the market will respond to a price increase of a certain size. If the monopolist predicts incorrectly and sets its price too high, the deadweight loss triangle will become larger and increase the social cost of the monopoly.

Finally, some economists have attempted to evaluate and quantify Learned Hand's dictum in the *Alcoa* case that monopoly "deadens initiative" and results in less efficient use of resources than would prevail in competitive markets. Monopolists may not have the same incentives to reduce costs; their managers may not operate under the same "crisis" conditions that affect competitors; they may become comfortable. Such phenomena undoubtedly exist in many firms. The extent to which they are more prevalent among monopolists than among competitors is difficult to quantify.[23]

13. In general, the utility will not innovate if any cost reductions or increased revenues are immediately passed on to customers. In most cases, however, the utility will be able to keep the increased profits, at least for a time, and this will give it some incentive to innovate.

14. See J. Ordover & R. Willig, An Economic Definition of Predation: Pricing and Product Innovation, 91 Yale L.J. 8 (1981).

15. For example, Berkey Photo, Inc. v. Eastman Kodak Co., 603 F.2d 263 (2d Cir.1979), cert. denied, 444 U.S. 1093, 100 S.Ct. 1061 (1980); California Computer Prod., Inc. v. IBM Corp., 613 F.2d 727 (9th Cir.1979). Both plaintiffs lost on the innovation issue. Claims of anticompetitive product innovation are discussed in § 7.8a.

16. See, e.g., C.R. Bard, Inc. v. M3 Sys., Inc., 157 F.3d 1340 (Fed.Cir.1998), in which a divided panel condemned a patent monopolist's reconfiguration of its tissue sampling machine so that it was compatible only with the defendant's disposable needles rather than those of others.

17. See J.K. Galbraith, American Capitalism 86 (Rev. Ed.1956).

18. United States v. American Can Co., 230 Fed. 859, 903 (D.Md.1916), appeal dismissed, 256 U.S. 706, 41 S.Ct. 624 (1921). See 11 Antitrust Law ¶ 1801a (1998).

19. United States v. Aluminum Co. of America (Alcoa), 148 F.2d 416, 427 (2d Cir.1945).

20. One study finding that firms become less efficient internally as the industries in which they operate become more oligopolistic is R. Caves & D. Barton, Efficiency in U.S. Manufacturing Industries (1990).

21. See ch. 8.

22. See ch. 18.

23. See H. Leibenstein, Allocative Efficiency vs. "X-Efficiency," 56 Amer.Econ.Rev. 392 (1966); L. De Alessi, Property Rights, Transaction Costs, and X-Efficiency: An Essay in Economic Theory, 73 Amer.Econ.Rev. 64 (1983). A good survey of the literature relating to productive inefficiency and monopoly is J. Siegfried & E. Wheeler, Cost Efficiency and Monopoly Power: A Survey, 21 Q.Rev. Econ. & Bus. 25 (1981).

1.3d. The Social Cost of Monopoly: Lost Competitor Investment

Figure 7 above also describes a third kind of welfare loss, denominated WL3. The WL3 rectangle is drawn away from the demand curve because it is an "externality"—something that shows up in neither the formation of the demand curve nor in the firm's calculation of its costs and profits. WL3's definition, existence or size is not clearly related to any of the cost or revenue functions that explain a firm's behavior.

Exclusionary practices, or rent-seeking, by the monopolist generally impose costs on the monopolist itself. The costs can be diagrammed, for their outer limit is determined by the wealth transfer, which is itself a function of the demand curve and the monopolist's marginal cost curve. A firm will not spend more in acquiring or maintaining a monopoly than the expected value of the monopoly. Thus the outer boundaries of monopoly rent-seeking are determined by the potential wealth transfer (WL2).

But monopoly rent-seeking also imposes inefficient losses on competitors or perhaps others, and these losses are potentially unlimited. They can certainly be larger than either the traditional deadweight loss (WL1) or the loss that results from rent-seeking (WL2). To take an extreme example, suppose that the world market contains two manufacturers of aircraft, each of which has a single plant. The CEO of one of the firms creates a monopoly by visiting the other firm's plant one night with a can of gasoline and a match, and burning it down. In this case WL1 is indeterminate, WL2 is the cost of the match, the gasoline, the opportunity cost of the CEO's time, and the risk and expected consequences of getting caught. At the very least, WL3 is the cost of the victim's destroyed plant, inventory and perhaps good-

will, of retraining employees whose jobs have been lost, and of reliance interests lost by broken contracts.[24]

What is the size of WL3 losses in real world monopolization or cartel cases? Generalizing is difficult, but it could be substantial.[25] Consider, for example, the Supreme Court's *Allied Tube* decision.[26] The plaintiff, Indian Head, had developed a plastic electrical conduit that threatened substantial injury to the market for traditional steel conduit. Defendant Allied, a manufacturer of steel conduit, conspired with others to "pack" a meeting of a standard setting organization with the result that approval of the plastic conduit was successfully delayed for several years. Because government building codes generally incorporated the organization's standards and numerous private contractors followed them voluntarily, the effect was that the plaintiff's plastic conduit could not be used in most construction.

In *Indian Head*, WL1 is the deadweight loss caused by any monopoly perpetuated by Allied's conduct.[26] WL2 includes the costs of packing the meeting and campaigning for disapproval of the plastic conduit, and the risk of a lawsuit and its costs. WL3 is the lost investment that accrued to Indian Head in research and development of a product that now has no market, or whose introduction into the market has been delayed. If Allied had succeeded in delaying plastic conduit indefinitely, Indian Head's entire investment in researching and developing a socially valuable product would have been lost.

Actually, the *Indian Head* situation may be a little more complex. Presumably, the demand curve for steel conduit would shift to the left in response to the introduction of Indian Head's product, which is a substitute for steel conduit. This would make steel conduit less profitable. The result of the conspiracy was to delay this shift, and this would yield a deadweight loss analogous to that caused by monopolization of a market in which no technological change was occurring.

Most bona fide monopolization cases produce substantial WL3 losses. Often the amount of WL3 loss will be proportional to the plausibility of the basic offense. For example, monopolizing conduct is most likely to succeed in markets where assets are specialized, durable and costly, because new entry into such markets can most easily be deterred. These markets are said to be subject to high barriers to entry.[27] WL3 loss is also most likely to be larger in such markets, because there is more likely to be investment that cannot be recovered in the event of failure. One important exception to this is strategic entry deterrence, or exclusionary conduct directed at potential, rather than actual, competitors. Potential competitors may be deterred easily precisely because they have not yet made irreversible investments in a market. WL3 losses in such situations are accordingly smaller.[28]

The model of WL3 losses limits the reach of arguments that antitrust should do away with competitor lawsuits, or at least severely circumscribe their role.[29] To be sure, most market injuries to competitors result from the increased efficiency of rivals—a theme to which this book often returns. Nevertheless, inefficient competitor injuries are real social costs and an important part of antitrust concern. Further, consumers are often not well positioned to redress these injuries, because they have inadequate information or face extremely difficult problems of organization or proof. (For example, the injured consumers in *Indian Head* are those that would have purchased plastic conduit but for Allied's antitrust violation.). By contrast, competitor injuries are often quite easy to quantify and known to competitors the instant they occur. This means that competitors may be in a better position to bring certain antitrust actions. They can sue earlier, when the social cost of inefficient monopoly is still rather small, and they may have a better knowledge base.[30]

The model of WL3 losses does suggest that the focus of the antitrust laws on lost profits in competitor suits is often misplaced. Lost profits are notoriously difficult to measure, especially when the plaintiff never had a chance to get into the market in the first place. Further, the real social burden of WL3 losses is lost *investment*—in the case of *Indian Head*, the resources spent in developing a product that cannot be marketed because of Allied's antitrust violation.

The question whether lost investments of this sort should form the basis of antitrust violations is controversial. First, because they do not show up in the "market," economists have been inclined not to calculate them as part of the social cost of monopoly. Second, natural free market forces, without the intervention of any anticompetitive practices, produce a great deal of lost investment. For example, much of the research and development engaged in by business firms fails to produce products that can be profitably produced.

But antitrust's concern is not with eliminating unproductive research and development, measured *ex post*. Rather, it is concerned with ensuring that, measured *ex ante*, the competitive incentives to R & D are maintained. For example, consider two firms racing to invent and patent a useable plastic conduit. First of all, a research joint venture might be a superior way to go about developing such a project, for it would entail one set of research expenditures rather than two. But our economy and the state of our legal policy is such that not every efficiency enhancing joint venture will be formed. As a result, two firms may be engaged in the highly inefficient activity of researching and developing the identical product. The winner gets a patent and twenty year monopoly, which will outlast the product's life; the loser gets nothing and its investment is lost. Presumably this loss is *not* a social cost of monopoly that the antitrust laws should take into account. Rather, it is a result of the kinds of inefficiencies that are an everyday occurrence in robustly competitive markets.

Suppose one of the firms wins the research race, not by doing better research, but rather by sabotaging the research of the rival, or perhaps by using ill-founded litigation strategically.[31] The differences between competitive behavior and noncompetitive behavior under such circumstances is that the competitive behavior (1) rewards the person who gets there first (and a product innovated today produces more social value than a product innovated tomorrow); and (2) the competitive behavior permits the *market* (or at least, the market as qualified by our patent laws) to determine whether there is room for both products or only one. By contrast, the anticompetitive behavior, such as sabotaging another's research, is calculated *ex ante* to yield the inferior solution. Normally, the person winning the patent race does not need to sabotage the person who is losing; it works the other way around. Considered *ex ante*, the monopoly created by the person who sabotages his competitor's research is not the kind of monopoly whose costs are offset by the increased incentive to research. Quite to the contrary. For policy reasons, then, we count this particular loss as a qualifying social cost that can raise antitrust's

24. For example, if a supplier has invested heavily in a contractual commitment to supply the victim firm with some input, that investment is now lost.

25. WL3 losses might also include practices that raise the marginal costs of rivals, thus causing deadweight losses in secondary markets. See I. Ayres, Rationalizing Antitrust Cluster Markets, 95 Yale L.J. 109, 117 n.42, Fig. 4 (1985); S. Salop & D. Scheffman, Raising Rivals' Costs, 73 Am. Econ. rev. 267 (1983).

26. Allied Tube & Conduit Corp. v. Indian Head, 486 U.S. 492, 108 S.Ct. 1931 (1988). A similar case is American Soc. of Mechanical Engineers v. Hydrolevel Corp., 456 U.S. 556, 102 S.Ct. 1935 (1982). For further discussion of *Indian Head*, see § 18.5; and see 13 Antitrust Law ¶¶ 2220, 2231 (1999).

27. See 2A Antitrust Law Ch. 4C (Rev. ed. 1995)

28. See S. Salop, Strategic Entry Deterrence, 69 Am. Econ. Rev. 335 (1979); O. Williamson, Predatory Pricing: A Strategic and Welfare Analysis, 87 Yale L.J. 284 (1977). For further analysis of WL3 and its implications for antitrust policy, see Hovenkamp, Antitrust's Protected Classes, note 5. For critiques, see W. Page, Optimal Antitrust Penalties and Competitors' Injury, 88 Mich.L.Rev. 2151 (1990); R. Markovits, Second Best Theory and the Standard Analysis of Monopoly–Rent–Seeking: A Generalizable Critique, a "Sociological" Account, and Some Illustrative Stories, 78 Iowa L.Rev. 327 (1993).

29. See E. Snyder & T. Kauper, Misuse of the Antitrust Laws: the Competitor Plaintiff, 90 Mich. L. Rev. 551 (1991). For a response, see R. Blair & W. Page, Controlling the Competitor Plaintiff in Antitrust Litigation, 91 Mich. L. Rev. 111 (1992).

30. See § 2.2c.

31. On the latter, see § 18.3b.

concern.[32]

One does not need to look at markets that are the subject of government intervention, such as the market for patents, in order to come up with analogous situations. Completely unregulated markets can produce a duplication of expenditures that might be regarded as a qualifying social cost when they are used for one purpose, but not when they are used for another. Consider the market for complex, high priced, and perhaps technically sophisticated structures. The developer who wishes to have such a structure may take competitive bids from intending builders. Looking *ex ante*, the cost of making a bid on a complicated project can be high—perhaps 2% or more of the product's final cost. Suppose that five bidders enter the contest, and the cost of making a bid is $100,000, but only one of the bidders can win. Further, the cost of making the bid is presumably sunk. A sunk cost is an investment that a firm will not be able to recover in the event of failure. In this case, the bid itself has no value to the loser. If all firms behave competitively, the process will yield a deadweight loss of $400,000 in bidding costs as compared with a process under which a single firm were asked to build the project and did so at the competitive price. The competitive bidding process is certainly wasteful of resources. Nonetheless, looking *ex ante* we can easily conclude that the bidding process is more efficient than any alternative mechanism for getting the project completed at a competitive price. We would not expect that the four losers would have a "damages action" against either the winning bidder or the developer.

Suppose, however, that four of the bidders had formed a cartel. When the fifth bidder refused to join in, the four undertook some exclusionary practice designed to make the fifth firm's bid unacceptable.[33] In this case the fifth firm would have an antitrust damages action.[34] Damages should be based on the fifth firm's lost investment—in this case, the $100,000 that the firm invested in a bid that it would have won, but for the cartel's exclusionary practice.

§ 1.4 Industrial Organization Theory and Economies of Scale

The field of economics known as industrial organization performs two important functions in antitrust analysis.[1] First, it can help us decide whether the perfect competition model is optimal for a particular market. Second, industrial organization can help us understand whether a particular firm's activities that affect market structure are efficient and should be encouraged, or inefficient and ought to be condemned. Indeed, the field of industrial organization developed in response to increasing policy concerns about the rise of "big business" in the late nineteenth century, and the resulting debate among lawyers concerning when antitrust condemnation is in order. Many of the basic doctrines of industrial organization theory were suggested first by lawyers in the context of litigation, and later adopted and formalized by economists.[2]

32. See H. Hovenkamp, Antitrust Policy and the Social Cost of Monopoly, 78 Iowa L. Rev. 371 (1993).

33. For example, the four might bribe a government official to refuse the fifth firm a license, or in the case of a public developer to reject the fifth firm's bid. They might also bring ill-founded litigation against the fifth firm, or bribe one or more of the fifth firm's employees to upset the bid; or they might agree with the fifth firm's suppliers to deny the fifth firm access to an essential input.

34. If the project were actually built at the higher bid price, the buyer of the project would also have a damage action for the monopoly overcharge.

§ 1.4

1. The classic text on industrial organization is E.A.G. Robinson's The Structure of Competitive Industry (rev. ed. 1958). A very useful and comprehensive contemporary text is F.M. Scherer & D. Ross, Industrial Market Structure and Economic Performance (3d ed. 1990). More technical accounts of the cutting edge of industrial organization theory are Handbook of Industrial Organization (R. Schmalensee & R. Willig, eds. 1989, 2 vols.); J. Tirole, The Theory of Industrial Organization (1988), which is particularly good on applications of game theory. A good beginners text is S. Martin, Industrial Economics: Economic Analysis and Public Policy (2d ed. 1994). A good economic introduction to the relationship between industrial organization and various regulatory and policy concerns is W. Kip Viscusi, John M. Vernon & Joseph E. Harrington, Jr., Economics of Regulation and Antitrust 166–179 (2d ed. 1995).

2. See H. Hovenkamp, Enterprise and American Law, 1836–1937 at chs. 22–25 (1991).

Many real world markets do not come very close to the classical model of perfect competition. In some markets this failure is an antitrust problem: the market would perform more efficiently if the firms behaved more competitively. In other markets, competition among large numbers of incumbents producing undifferentiated products is simply not the optimal structure.

1.4a. The General Case of Economies of Scale

The single largest factor tending to undermine perfect competition is economies of scale. The model of perfect competition is premised on the notion of a market containing many equally efficient firms, each indifferent to the output decisions of others. Within this model, firm size is not a factor in competitor decisions, because the model assumes constant returns to scale: production and distribution costs do not vary with size. A small firm can thus compete quite effectively with a large one. Suppose, however, that one firm develops a new process that enables it to produce the product at substantially lower cost. In order to take advantage of this new process, however, the firm must build a plant capable of serving one half of the existing market. Now incumbent firms can no longer be indifferent to the price and output decisions of the innovator.

Most economies of scale are not as dramatic as the illustration suggests. However, economies of scale obtain in most industries, and they range from the trivial to the very substantial. Technically, an economy of scale exists whenever the costs per unit of some input decrease as volume increases.[3] The following examples illustrate the manifold presence of scale economies in a wide variety of industries.

1) To drive a truck from point A to point B costs $100, whether the truck is full or half empty. As a result, the full truck can transport its cargo at a lower cost per pound than the half empty truck.

2) A 30–second television commercial advertising automobiles costs $100,000, whether the manufacturer has 4,000 dealerships across the country and produces 10,000,000 cars per year, or has 300 dealerships and produces 90,000 cars per year.

3) To set up an automatic metal lathe to turn out a particular machine part costs $100 in labor. Once the lathe is set up, the costs of turning out the parts is $1.00 each. If the lathe is set up to turn out a single part, its cost will be $101.00. If the lathe is set up to turn out 10,000 parts, their cost will be $1.01 each.

4) A procurement department and legal staff spend $2000 to negotiate and draft a contract to purchase an essential raw material, whether the company is buying 50 units of the material or 5000 units.

5) A manufacturer of essential medical or industrial supplies must always keep one production machine in reserve, so that a breakdown will not interrupt production. If it produces with a single machine operating at a time, it must therefore maintain capacity equal to twice its actual output. If it produces with eight machines, however, it needs to maintain only nine machines, a capacity equal to 12% more than output.

6) A production process requires 40 discrete functions. If a firm has ten employees, each must perform, on average, four different functions. If the firm has 4000 employees, no single employee will have to perform more than one function, in which she will be a specialist. If she becomes ill, another specialist in the same function will replace her.

7) The transaction costs of borrowing money (or raising equity capital) are 2% for blocks of $1,000,000; 1% for blocks of $10,000,000; or .5% for blocks of $100,000,000.

8) The development of a new manufacturing process reduces the cost of manufacturing widgets by 50¢ per unit. The research and development costs for inventing the new pro-

3. By contrast, an economy of *scope* exists when there are economies to performing two different economic activities at the same time. For example, if corn starch and corn flakes are separate products from a kernel of corn, a company that produced both products simultaneously in the same plant might have lower costs than individual companies that produced each in separate plants. For a brief technical discussion, see J. Panzar & R. Willig, Economies of Scope, 71 Am. Econ. Rev. Pap. & Proc. 268 (1981).