Antitrust Law

Economic Theory and Common Law Evolution

KEITH N. HYLTON
Boston University

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2003

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times Ten 10/13 pt. *System* QuarkXPress [BTS]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Hylton, Keith N.
Antitrust law : economic theory and common law / Keith N. Hylton.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-79031-X (hardback) – ISBN 0-521-79378-5 (pbk.)
1. Antitrust law – United States. 2. Antitrust law – Economic aspects. I. Title.

KF1649 .H95 2003
343.73′0721–dc21 2002017406

ISBN 0 521 79031 X hardback
ISBN 0 521 79378 5 paperback

Contents

I

Economics

In this chapter, I will introduce economic concepts that I plan to use throughout the text, set out the basic economics of monopoly, and compare monopoly with its polar opposite, perfect competition. I also will discuss some relatively new topics, such as transaction-cost and information economics, and their relevance to antitrust policy.

I. DEFINITIONS

A. Monopoly

A monopolist is a single supplier of a good. However, this definition is too simple, because it includes firms that become dominant by being the lowest-cost competitor and those that obtain an exclusive franchise from the state. As far as antitrust policy is concerned, there is a big difference between these two cases. Because the simple, "single-supplier" definition is potentially misleading, one should focus on market conditions. The crucial feature of monopoly status is the *absence of competition* from other firms.

The common example of monopoly in our lives is local telephone service, provided in most places in the United States by a regulated, privately owned monopoly. However, even here competition from wireless and optical fiber companies has eroded the monopoly status of the local telephone companies. One of the purest monopolies in recent memory was Aeroflot, the airline of the former Soviet Union. Before the breakup of the Soviet Union, there were no competing airlines.

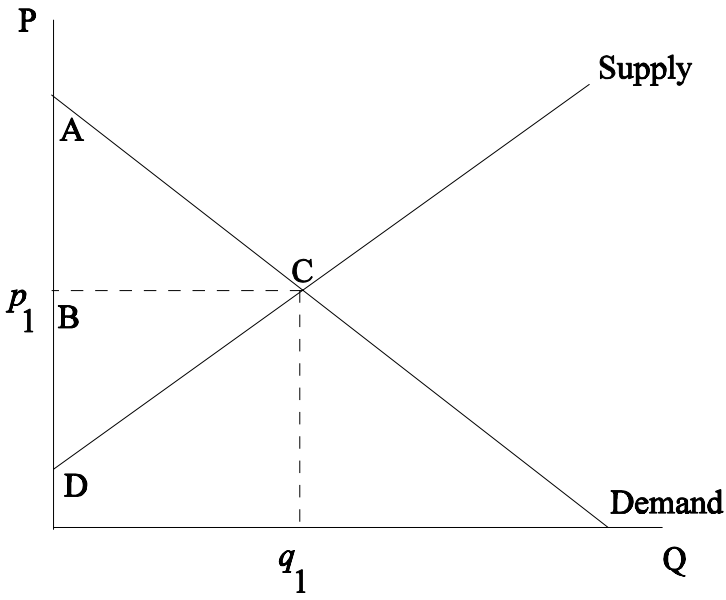

Figure 1.1

B. Market Price

A *market equilibrium*, where the quantity demanded by consumers equals the quantity supplied by producers, generates a market price, as shown in Figure 1.1. The downward sloping line is the demand curve and the upward sloping line is the supply curve. Think of the demand curve as a schedule of bids offered by consumers. Each point along the curve is a maximum price that at least one consumer is willing to pay. The horizontal axis measures the total quantity demanded at a given price, and since each consumer would accept the item at a lower price, quantity demanded increases as price falls. Similarly, one can think of the supply curve as a schedule of minimum asking prices stated by producers. Since each producer is willing to sell the good at a price at or above his asking price, the total quantity offered for sale at a given market price (measured by the horizontal axis) increases as price rises.[1]

[1] More technically oriented treatments typically explain that each consumer has a schedule of bids for each quantity desired. Consumers offer less per unit for higher quantities because the utility gained per unit of consumption falls as consumption expands. The market demand curve is the "horizontal sum" of the individual demand schedules. The student trained in economics may prefer to think in these terms. I have attempted to simplify the presentation in the text.

In a market with many producers and consumers, none of them actively sets the equilibrium price. It is, in a passive sense, determined by the actions of the marginal consumer and marginal producer. The marginal consumer (point C in Figure 1.1) is just indifferent between buying the good and going without it, given the market price. Similarly, the marginal producer (also at C) is indifferent between selling at the market price and keeping his output. The inframarginal consumer (to the left of C along the demand curve in Figure 1.1) is willing to pay more for the good than is the marginal consumer, and the inframarginal producer (to the left of C along the supply curve) is willing to part with the good for a lower price than the marginal producer would accept. The price in an exchange between inframarginal actors is indeterminate – it is any level between the maximum the consumer is willing to pay and the minimum the producer is willing to accept.

To see the role played by marginal actors in the determination of equilibrium, suppose the price is initially set above the level that equalizes the amounts demanded and supplied. Suppliers would offer a quantity larger than consumers were willing to purchase, and as a result some sellers would be unable to find buyers. Among them would be inframarginal sellers, who would cut their asking prices in order to make a sale. This process would continue until the equilibrium price is reached.

C. Market's Contribution to Wealth

Because the marginal consumer determines price, all other consumers (inframarginal) gain by making trades in the market. *Consumers' surplus* measures the gain to consumers from taking advantage of the market: some consumers would still buy the good at a higher price, but they can purchase it at the cheaper market price. Similarly, because the marginal producer determines price, *producers' surplus* measures the gain to producers generated by market transactions.

The diagram in Figure 1.1 also illustrates the incremental wealth generated by the market, which is equal to the sum of consumers' surplus and producers' surplus. Consumers' surplus is the area ABC, and producers' surplus is the area DBC. The total surplus, or the market's contribution to wealth, is maximized when price is equal to the market equilibrium level p_1 and quantity is equal to the market equilibrium level q_1. Because total surplus hits its maximum at the market equilibrium, I will refer to this as the *social optimum*.

Although this may seem an unusual way to measure incremental wealth,[2] this is the approach Adam Smith emphasized in arguing against the mercantilist policies followed by England and other European countries over the eighteenth century.[3] The doctrine of mercantilism, still with us today in many quarters, held that a government should manage foreign trade in order to maximize gold reserves. To the mercantilists, this was how a country became wealthy. In practice, the doctrine necessitated a strategy of maximizing exports and minimizing imports. Adam Smith's argument, startlingly counterintuitive at the time and still misunderstood by the majority of governments today, was that the mercantilists' conception of wealth was invalid and that their policies were likely to reduce rather than increase wealth. A market's real contribution to wealth is the difference between the value of the benefits a good provides and the resource cost of its production. Smith argued that a policy of free trade in competitive markets maximizes this measure of incremental wealth. Of course, understanding Smith's argument requires some familiarity with the properties of competitive markets. I take up that topic next.

D. Defining Perfect Competition

A competitive equilibrium satisfies the assumptions of the model of perfect competition, which are as follows.

1. Atomism. The output of each seller and the consumption amount of each buyer is a small fraction of the total output of the market, so no buyer or seller can have more than a very small influence on market price or quantity. Alternatively, each buyer and seller takes market price as given.

We could speak generally of a spectrum with atomism on one end and monopoly on the other. Of course, the theoretical endpoints are hardly ever observed. Atomism, in its extreme version, requires an infinite number of infinitesimally small producers and consumers. Monopoly requires a single seller, but even where we do find a single seller of an item, often suppliers of close substitutes constrain the monopolist's price-setting decisions.

[2] It is important to note the difference between stocks and flows. Consumers' surplus is a flow while wealth is a stock. In view of this, I have referred to consumers' surplus as a component of incremental wealth.

[3] *An Inquiry into the Nature and Causes of the Wealth of Nations* (Edwin Cannan, ed., New York: Modern Library 1994).

2. Perfect Information. Consumers can distinguish between different goods. They also know if one seller is offering a particular good at a lower price than another seller. Really, all we need is that information must get around reasonably fast. The assumption of perfect information simplifies the matter.

Obviously, the assumption is not an accurate description of the world.[4] In the real world, we see firms advertising. We could make the model resemble the real world more closely by assuming that information is a commodity that must be supplied.

However, once we assume information must be supplied, the reason for making the simplifying assumption of perfect information starts to become clear. The market for information is peculiar. Information is a *public good*, in the sense that a decision to supply it to one person generally means that the good is also available to others. For example, the purchaser of a newspaper may read it and then give it to a friend. Because the information can be shared, the producer may not receive compensation for the benefits conferred upon a large number of consumers, and in this case the market may provide insufficient incentives to produce news. This is illustrated in Figure 1.2. The forward-shifted demand curve includes the benefits of newspapers to nonpaying readers. The social optimum is at (p_2, q_2) rather than the market equilibrium (p_1, q_1).

This example suggests that relaxing the assumption of perfect information immediately introduces some element of market failure into the model. Consider the case of advertising to inform consumers of the existence of a better mousetrap. Suppose there are competing sellers of this new mousetrap. A seller who advertises the mousetrap cannot be sure that the benefit will accrue to himself alone, because he cannot limit the message only to consumers who will purchase from him. Since some of the benefits may go to other sellers, his incentive to pay for informative advertising is attenuated.

In the extreme case, failure of the assumption of perfect information can make a market virtually infeasible. The best example is the problem of *adverse selection* in the insurance market. Suppose there is a

[4] A related and more fundamental criticism is that the assumption of perfect information ignores the central problem that needs to be explained: how privately held information is revealed and communicated among market participants, see F. A. Hayek, *Individualism and Economic Order* 77–106 (London: Routledge & Kegan Paul, 1949). Although Hayek's critique has important implications for antitrust policy, space will not permit me to cover it here.

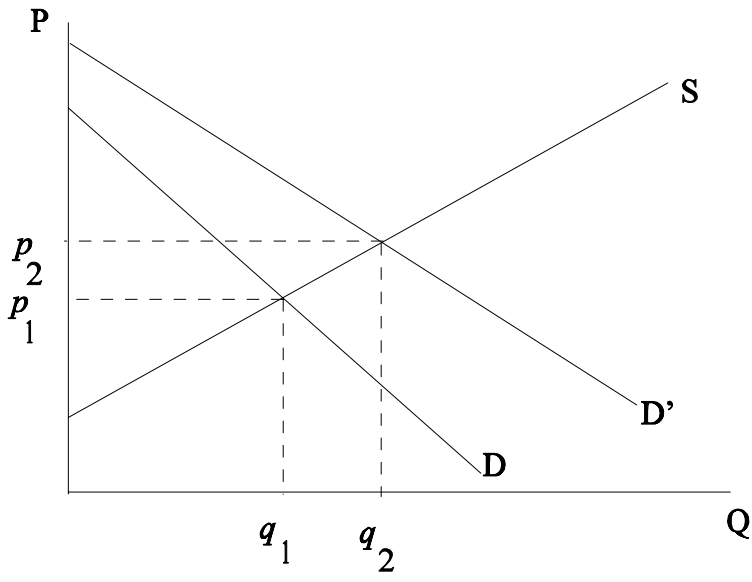

Figure 1.2

continuum of risk levels among potential insurance purchasers, and the insurer cannot determine the risk level of each applicant. The insurer's price will be a weighted average of the prices that should be charged to each type, the weights reflecting the anticipated shares of each risk type in the insured population. If some relatively low-risk customers exit the relationship and insure themselves or do without insurance altogether, then the price must be increased for the relatively high-risk customers who remain. But this may lead others to drop their policies, and so on. In the end, only the most risky customers seek insurance, and with little to be gained from pooling their risks, the market vanishes.

3. Mobility. Resources flow easily from one market or sector of the economy to another: no barriers to entry exist. Without mobility, monopoly power becomes possible. Simply meeting the assumption of atomism does not eliminate the possibility of market power.

To take a concrete example, consider the market for attorneys. In the United States, there are too many of them to count. It would seem, therefore, that the atomism requirement is satisfied. However, the market is not perfectly competitive because not everyone who could perform as an attorney is permitted to enter the market. Every attorney must pass

a bar exam and be sworn in to the state in which he or she wishes to practice. The bar passage requirement reduces the total number of attorneys and allows them to earn a return in excess of the opportunity costs of the skills and resources employed in that profession.

4. No Third-Party Effects. The model of perfect competition assumes there are no *externalities*, that is, third-party effects. The parties who contract over the supply of a good or service bear all of the costs and benefits associated with the production of that good or service.

Externalities lead to production levels that deviate from the social optimum. For example, consider the case of a company that produces chemicals and also pollutes the water as a byproduct. The company produces too much from society's point of view. The total cost of the company's output is more than the production cost borne by the company, it also includes the costs generated by the pollution. If the company were forced to bear the pollution costs, it would demand a higher price in order to supply the market. Put another way, the supply curve for the chemical producer would shift back, as shown in Figure 1.3, reflecting the higher price demanded for each level of output. The

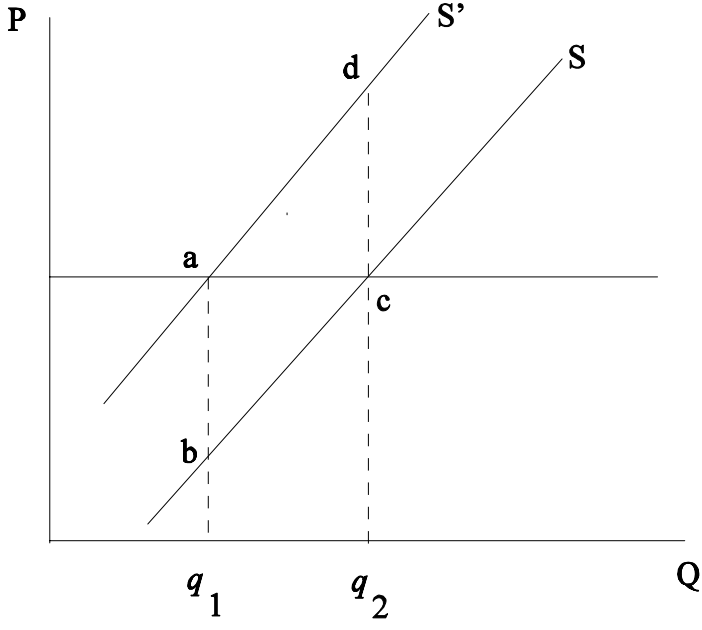

Figure 1.3

upward-shifted supply curve in Figure 1.3 reflects the real costs of producing chemicals. As the diagram also shows, given any market price, the firm overproduces, relative to the social optimum (which is at q_1).

One way to correct the overproduction demonstrated in this example is to tax the chemical producer. The company's supply should be reduced by taxing it at a rate that reflects the costs generated by the pollution it imposes on society.

A more general approach to solving this problem was suggested by Ronald Coase.[5] Coase demonstrated that in a regime in which transaction costs were zero, and property rights well-defined, resource allocation would be efficient. To see why this holds, consider again the example of the chemical producer who pollutes the water. Suppose a downstream firm finds that it must clean the water in order to use it in production. If it is less costly for the upstream chemical producer to reduce its production than for the downstream firm to clean the water, then the downstream firm will have an incentive to offer a payment to the upstream chemical producer in exchange for a reduction in the upstream producer's level of output. The incentive for such a side payment remains as long as the gain from cleaner water to the downstream firm (area **abcd** in Figure 1.3) exceeds the loss from cutting back production to the downstream firm (area **abc**). As Figure 1.3 suggests, the side payments will continue until the upstream producer cuts back to the optimal level q_1.

5. Homogeneous Product. Products are not differentiated. For example, a seller of wheat really sells standard wheat – nothing fancier or different from what every other wheat seller offers.

This assumption implies that markets cannot be divided up into small enough portions to violate the atomism assumption. If, for example, the market for wheat could be divided into one million markets for different types of wheat, one of those one million markets could likely contain only one firm. Thus, the homogeneity assumption provides another way of avoiding monopoly.

Homogeneity also helps avoid the informational problems suggested above. Suppose there were several brands of wheat and consumers could not distinguish one from another. Then an inferior brand might sell for the same price as a superior brand, because consumers were unable to make fully informed choices.

[5] R. H. Coase, *The Problem of Social Cost*, 3 J. Law & Econ. 1 (1960).

E. Economic Profit

Economic profit is the excess of revenue over costs, where costs include compensation for risk-taking and the opportunity cost of capital. This is not the same as accounting profit, which makes no attempt to include risk-taking and lost opportunities as elements of total cost. A firm may be earning positive accounting profits and negative economic profits. This is why one cannot infer monopoly power simply from observing the profit reports of a company.

A simple story illustrates the concept of economic profit. Suppose a wealthy ice cream lover donates two plots of land to a company that runs a chain of ice cream parlors. One plot is in Quiet Square, a sleepy, small-town intersection that rarely sees crowds. The other plot is on Busy Street, smack in the middle of downtown Busy City, an area full of pedestrians from sunrise to sunset. One would not be surprised to find that the Busy Street parlor makes a substantially greater accounting profit than the Quiet Square parlor. However, the relation between their economic profits may be the opposite. To measure the economic profit of the Busy Street parlor, one must subtract from accounting profit an estimate of the rental value, or opportunity cost, of the plot of land on Busy Street. Economic profit at either ice cream store is measured by the extent to which accounting profit exceeds the rental price for the location.

II. PERFECT COMPETITION VERSUS MONOPOLY

A. Perfect Competition

The fundamental result of the model of perfect competition is the following: *In long run competitive equilibrium, firms earn zero economic profits.*

This happens because of entry and exit. If firms earn positive economic profits, then rivals will enter the market. Entry continues until the increase in supply pushes price down to a level that just compensates for the cost of producing and the opportunity cost of capital and managerial skill. If firms earn negative economic profits, exit will occur until economic profits return to zero.

It is important to keep in mind that entry and exit occur in response to economic profits, not accounting profits. Second, economic profits go to zero in the long run, not the short run. Nothing in the model of perfect

competition suggests that firms cannot earn economic profits in the short run. Indeed, the appearance of economic profits (economic losses) in the short run causes entry (exit).

Although the five assumptions of perfect competition described in the previous section should be sufficient to generate the zero economic profits proposition, an intermediate set of assumptions (almost all of them derivable from the initial five) are useful in analyzing the long run equilibrium of a perfectly competitive economy.

The first intermediate assumption is that each individual firm faces an infinitely elastic demand curve. The *elasticity of demand* measures the responsiveness of the quantity demanded to the changes in the price of the good. A zero demand elasticity means that a price change has no effect on the quantity demanded. Infinite elasticity means that a firm can produce as much as it wants to sell at the equilibrium price without that increase in quantity supplied having any effect on the market price.

Because of this assumption, the firm in a competitive economy becomes a "price taker," that is, it takes the market price as given – fixed, not subject to its influence. A firm can certainly charge a price different from the market price; however, the assumption implies that the firm has no incentive to do so. Suppose the firm sets its price above the competitive level. It will sell nothing, because consumers can buy at the market price from another firm. Suppose the firm sets its price below the competitive level. Then it sells the same amount as it would at the competitive price, but it will make less revenue because it sold at a lower price.

The second intermediate assumption is profit maximization. In long run competitive equilibrium, economic profits are zero, which implies that price is equal to average cost. Let C = production cost, AC = average cost (C/q), MC = marginal (or incremental) cost. Then profit = $pq - C = q(p - AC)$, so positive profit implies $p > AC$, and zero profit implies $p = AC$. Since the firm is also maximizing profits, price must equal marginal cost ($p = MC$). Why? The firm is maximizing profits, which means it increases output until marginal revenue equals marginal cost ($MR = MC$), or that it will produce each unit that brings in as much or more revenue than it costs. When the demand curve is infinitely elastic, $MR = p$. Thus, in long run competitive equilibrium $MR = MC = AC = p$.

Profit-maximization is not a strong behavioral assumption because a competitive environment more or less forces firms to maximize profits. Suppose a firm chose not to maximize profits. Since economic profits, among profit-maximizing firms, are zero in the long run, a firm that did not maximize profits would earn a negative economic profit. The owners

of the firm would then come under pressure to sell the assets or transfer them to some other use.

The third intermediate assumption is that the individual firm faces a U-shaped long run average cost curve. It has this shape because of increased opportunities for specialization (as scale increases) and substitution of more productive inputs. Consequently, we can view the firm as initially drawing on factors of production uniquely suited to the firm's activity. Because these specialized factors are unusually productive, they drive down long run average cost. To see this, let L represent the only (variable) input and let its price be w. Also, let AP_L represent the average product of L. In the long run (where all factors are variable), $AC = C/Q = wL/Q = w/AP_L$. Thus as average product increases, average cost falls. Later, as the gains from specialization are exhausted and the firm begins to draw on factors that are not so well suited, the long run average cost curve begins to rise.

I will not present a detailed discussion of the relationship between short- and long-run cost curves, but I will note here that short-run cost curves typically are U-shaped for different reasons. Short-run cost curves fall initially because of fixed costs, and, in some cases, the rising average product of the variable factor; and begin to rise at some point because of diminishing returns. With one factor of production fixed, the variable factor becomes less productive as output expands.

On the basis of these assumptions, a simple diagram illustrating the process of competition is shown in Figure 1.4. Suppose price rises above

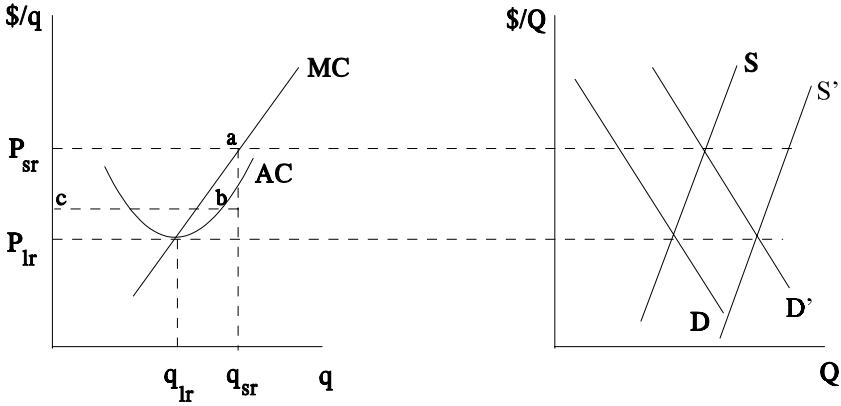

Figure 1.4 P_{sr} = short-run price
P_{1r} = long-run equlibrium price
Area P_{sr} abc = short-run profit

the competitive level (because demand shifts outward from D to D'). Firms expand output, and see profits in the short run. The profits provide a signal that leads to entry. Entry causes the industry supply curve (see S) to shift outward (S'), until price returns to the long-run equilibrium. At that level, price equals the minimum of the long-run average cost curve, which means that firms produce goods in a method that economizes on production resources.

The long-run competitive equilibrium is *efficient* in the following sense. Total welfare is maximized because price, which measures the marginal benefit to consumers, equals marginal cost. Thus, no alternative price-output combination could increase society's welfare relative to the long-run competitive equilibrium. Furthermore, because price is equal to the minimum of the long run average cost curve, the long run competitive equilibrium achieves this welfare-maximizing price-output combination in a manner that economizes on productive resources.

B. Monopoly

1. Basics. The monopolist, unlike the perfectly competitive firm, does not face an infinitely elastic demand curve. This is because the monopolist's demand curve *is* the market demand curve. Like the competitive firm, the monopolist tries to maximize profit. To do this, the monopolist expands output until the increase in revenue attributable to the last unit just equals the incremental increase in cost ($MR = MC$).

In order to achieve this, the monopolist restricts output and charges a higher price than would a firm in a competitive industry operating under the same cost conditions. This occurs because marginal revenue for the monopolist is always less than what it would be for a competitive firm facing the same price. Why? Because an increase in output raises total revenue by price times the increase in quantity, but also reduces per unit revenue by causing a decrease in the price. Thus, under monopoly $MR < p$.

The results under monopoly are the following. (1) Economic profits are positive ($p > AC$). (2) Part of the wealth of consumers is transferred to the monopolist (see the rectangle P_mABC in Figure 1.5). (3) Part of society's wealth is wasted. This is illustrated in Figure 1.5, which assumes a horizontal average cost curve for simplicity. The area labeled "deadweight loss" (triangle **ABD**) measures waste. This waste of society's wealth occurs because consumers are willing to pay a price that exceeds

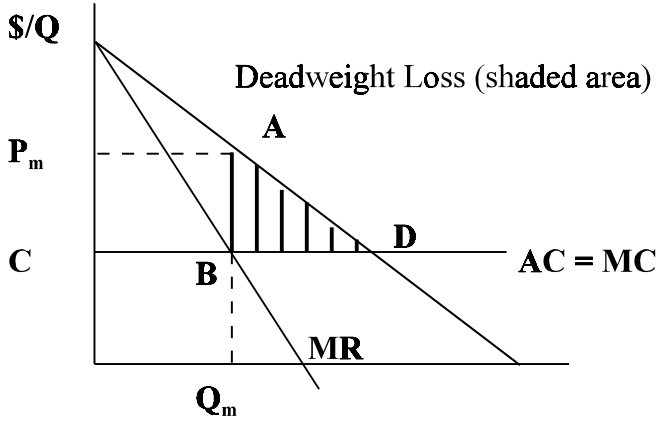

Figure 1.5 P_m = monopoly price
Q_m = monopoly output
Area P_mABC = monopoly profit = monopoly transfer

the marginal cost of producing additional output, but the monopolist does not supply the additional output. Thus, the deadweight loss triangle contains the set of potential welfare-enhancing exchanges that are forgone by the monopolist.

The transfer of wealth shown in Figure 1.5 does not always happen as described. Sometimes increasing costs consume all of the wealth: it is still a transfer, but not to the monopolist's profits. For example, the transfer may go to the owner of the assets needed to acquire the monopoly. Consider the case of taxicab medallions in New York City. The last report of a sale of a taxi medallion in New York City listed the price as $182,000,[6] which suggests buyers expect to earn monopoly rents.

Medallions are an example of a *rent-seeking* expenditure: an investment that does not generate wealth for society. In the case of monopoly, people bid for rights to a portion of wealth transferred from consumers. After acquiring the right, they need to earn the rent to pay off the debt the acquisition generates. In some cases, the rent may just meet the debt payments. Indeed, it follows from the model of perfect competition that if the market for acquisition of monopoly status is competitive, rents earned by winners will be merely sufficient to cover the acquisition costs. Put another way, if the market for acquisition of monopoly status is

[6] See Sheryl Fragin, *New York's Terror Taxis, Explained*, The New York Times, Sunday, August 21, 1994, page 9, Sunday Business Section.

competitive, we should observe a phenomenon that can be described as *ex ante rent dissipation*.

Sometimes, managers and other agents of the firm transfer the monopoly rents into production costs – a process that can be labeled *ex post rent dissipation*. We just considered the example of the taxicab medallion owner who buys into a monopoly and needs to earn the stream of monopoly rents in order to pay off the debt (an example of ex ante rent exhaustion). But suppose the owner already has the medallion, and a new law limits the supply of medallions, and suppose further that price regulation discourages price competition among the medallion owners. The owners may find other ways to compete. For example, they may purchase fancy cars to use as cabs, or offer drinks or food to customers. This type of competition occurred in the airline industry under regulation by the CAB. Airlines competed in nonprice categories, and the competition drove service costs up.[7] Unions demanded a share of the rents and managers paid them off to avoid labor problems. This process converted monopoly rents over time into costs.

Recall that the waste of society's wealth shown in Figure 1.5 results because the monopolist forgoes several wealth-enhancing trades. Both sides could gain if a transaction took place within the deadweight loss triangle. The portion of the demand schedule between points A and D in Figure 1.5 shows the maximum prices consumers are willing to pay, which reflect their valuations of the benefits they derive from the monopolist's product. The marginal cost the seller incurs by supplying additional output is given by the portion of the marginal cost curve between points B and D. Since the consumer's willingness to pay equals or exceeds marginal cost at all output levels along these segments, both sides would gain at any transaction price less than the consumer's maximum offer price and greater than marginal cost.

Why doesn't the monopolist supply the additional output? To do so while still earning at least the monopoly profit would require the firm to *price discriminate* by setting a lower price for consumers whose maximum offer prices are on the portion of the demand curve between points A and D in Figure 1.5. If the monopolist could perfectly price-discriminate, by charging each consumer a price equal to the consumer's maximum offer price, then the monopolist would expand output up to the competitive level (point D). Note that this implies that there would

[7] See, for example, Richard A. Posner, *The Social Cost of Monopoly and Regulation*, 83 J. Pol. Econ. 807–27 (1975).

be no deadweight loss: the monopolist supplies the competitive level of output, and takes all of the additional surplus in the form of profit. In the more realistic scenario in which the monopolist cannot charge each new consumer a price equal to that consumer's maximum offer price, the monopolist still has an incentive to expand output if he can sell additional output to the consumers between points A and D on the demand curve at some price between those two points (and note that unlike the perfect price-discrimination case, the new consumers will gain also). However, the monopolist will not go all the way to the competitive level of output in this case.

The problem with price discrimination is this: how would the monopolist prevent the low-offer-price consumers from reselling to other (high-offer-price) consumers? The administrative costs of setting up and enforcing a price discrimination scheme (in which a monopolist charges different prices, unrelated to the costs of supplying those units, to different consumers) could outweigh the additional benefits to the monopolist. The standard analysis of monopoly implicitly assumes the administrative costs of price discrimination exceed the benefits to the monopolist.

2. Stability of Monopoly. Because monopolists earn profits in excess of opportunity costs, they attract entry. It follows that for monopolies to continue, barriers must prevent entry by competitors. But what is a barrier to entry? Is having to build a plant a barrier to entry?

Generally the literature identifies two types of entry barriers: *natural* and *artificial*. Let us start with artificial barriers to entry. There are two kinds of artificial barriers: *government created* and *privately created*. Government created artificial barriers can include: (1) patents, (2) taxicab medallions, (3) government franchises (e.g., electricity supply) or exclusive contracts, and (4) licensing. Examples of privately created artificial barriers are: product differentiation, advertising, exclusivity contracts, and product tying.

The government created barriers are fairly easy to understand, but the theory of private barriers poses some difficulty. Product differentiation sometimes acts as a barrier to entry because it creates brand loyalty and therefore makes it more difficult for a rival to enter and compete for consumers. Exclusive dealing contracts create barriers by foreclosing the market to rivals. For example, if firm A has an exclusivity contract with the only supplier of a vital input, it would be difficult for a rival to enter and compete against firm A. Product tying also tends to exclude

rivals by forcing them to enter at two levels (the tying and tied product) in order to compete against the seller of the bundled product.

The theory of privately created barriers is a field of controversy in antitrust policy.[8] On one extreme is the expansive view suggested in the work of Joe Bain,[9] and on the other the view, suggested in the work of Harold Demsetz,[10] that the government creates the only real entry barriers – and even then the concept is troubling in Demsetz's view because government necessarily plays an important role in defining property rights. George Stigler[11] took an intermediate position, labeling such things as product differentiation a barrier to entry only if the cost of differentiating a product is higher for an entrant than an incumbent firm.

Admittedly, the private barrier theory can go to an absurd extreme. Consider the necessity of building a plant in order to produce the good. Is this also a privately created entry barrier? If building a plant is not a privately created entry barrier, then why is it a privately created barrier to form an exclusive dealing arrangement with the only supplier of a vital input? One could say that in the former case, any potential rival could build his own plant, while the latter example involves an exclusive arrangement that cannot be duplicated. But suppose the incumbent firm's plant uses up the best location available, and suppose there are substitutes to the vital input?

Natural barriers make up the second general class of entry barriers. The classic example is that of an "increasing returns" or "high fixed costs" monopoly, such as railroads, electricity suppliers, and water suppliers. The phenomenon of a falling average cost through the scale of production appears in each of these examples. Figure 1.6 illustrates the cost curves for a high-fixed-costs monopoly. The marginal cost curve lies below the average cost curve at all output levels. Because of this, the firm has an incentive to expand output even when it would not be able to cover all of its costs at the higher output level.

If the average cost curve declines throughout, as shown in Figure 1.6, then competition will result in leaving one firm in the industry. That one firm will have a *natural monopoly*. It would be inefficient to have two

[8] For an illuminating discussion of the problems of definition in this area, see Harold Demsetz, *Barriers to Entry*, 72 American Economic Review 47–57 (March 1982).

[9] Joe S. Bain, *Barriers to Entry* (Cambridge, MA: Harvard University Press, 1956).

[10] Demsetz, *supra* note 8.

[11] George J. Stigler, *The Organization of Industry* 67–70 (Chicago and London: University of Chicago Press, 1968).

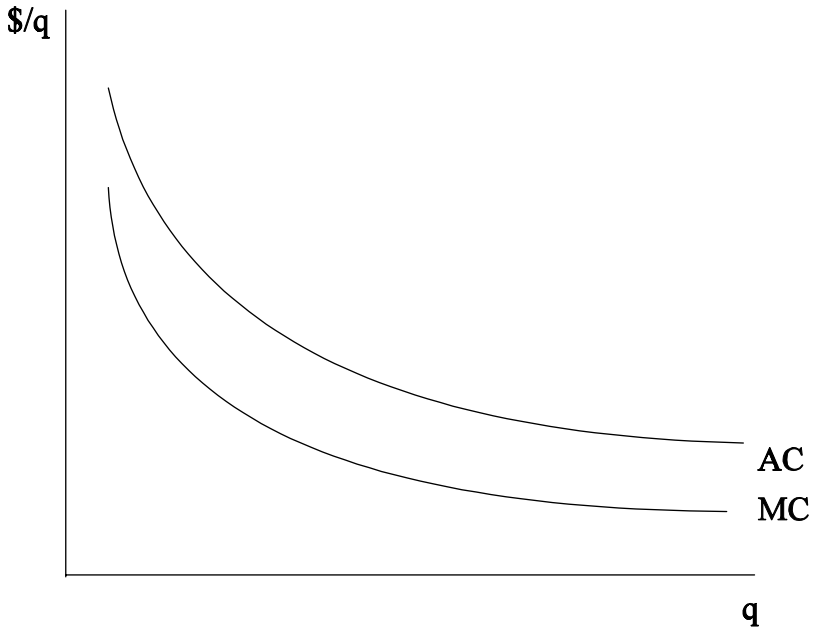

Figure 1.6

firms supplying the good. Why, for example, should two firms run sewer pipes, or telephone wires through the same portion of a city?

In many of the early Sherman Act cases, railroads argued that they suffered from the declining average costs phenomenon, and should therefore be allowed to enter into cartels.[12] Unrestrained competition, they argued, would lead to ruinous competition: in the end, only one railroad would survive, which would then charge a monopoly price.

Economists today generally do not accept this argument, and antitrust courts at the turn of the century rejected it. As stated, it is a weak argument. The fundamental problem is that average costs are unlikely to fall through the entire range of production. At some point, it would become difficult to gain efficiencies by simply expanding. No individual railroad ever grew so large as to test this proposition, but it is implausible anyway.

In spite of the implausibility of the ancient ruinous competition argument, new research has suggested that the argument for price-fixing

[12] See, for example, *U.S. v. Trans-Missouri Freight Assn.*, 166 U.S. 290 (1897).

cartels may have had some validity. Indeed, the improbability of a firm experiencing declining average costs through the entire scale of production seems now, in light of recent research, to provide additional support to the policy advanced by the railroads. Chapters 4 and 5 will discuss this subject.

3. Deadweight Loss: Large or Small? There is a strong argument for the position that the deadweight loss due to monopoly is small, too small, in fact, to concern enforcement authorities. In 1954, Arnold Harberger estimated that deadweight loss triangles distributed across the economy add up to at most 1/10 of 1 percent of national income.[13] Harberger's estimate touched off several attempts to independently estimate the size of the welfare loss, most of which reached conclusions similar to his and a few that generated numbers as high as 4 to 7 percent.[14] Although Harberger's analysis remains a matter of contention, the clear implication is that the benefits of anti-monopolization efforts are small, and probably less than the costs of legislative and enforcement activity directed at the monopolization problem.

The view that deadweight loss is large has received its best theoretical support from the literature on rent-seeking, which began in 1967 with an article by Gordon Tullock.[15] Tullock made the point that the expectation of earning profits in excess of opportunity costs would generate efforts to gain entry or ownership of the right to monopolize. In a fully competitive ex ante market in monopolization rights, the winning bid would equal the expected flow of monopoly rents. The resulting winner

[13] Arnold C. Harberger, *Monopoly and Resource Allocation*, 44 American Economic Review 77–87 (May 1954). Harberger's approach tried to err on the side of overestimating rather than underestimating the resource misallocation due to monopoly power. One might argue that the effort was flawed from the start because the size of the misallocation was itself a reflection of the threat of antitrust enforcement in the period of Harberger's sample, 1924–8. But it is unlikely that antitrust enforcement had a large impact on the size distribution of firms over the 1920s, see George J. Stigler, *The Economic Effects of the Antitrust Laws*, 9 J. Law & Econ. 225–58 (October 1966). Furthermore, there is evidence that early antitrust enforcement, by outlawing cartels and leaving mergers unregulated, may have caused an increase in market concentration levels, see George Bittlingmayer, *Did Antitrust Policy Cause the Great Merger Wave?*, 28 J. Law & Econ. 77–118 (April 1985).

[14] For an excellent summary, see F. M. Scherer, *Industrial Market Structure and Economic Performance* 459–62 (Houghton Mifflin, 2d ed. 1980).

[15] Gordon Tullock, *The Welfare Costs of Tariffs, Monopolies and Theft*, 5 Western Economic Journal 224–32 (1967). For a survey of the literature on rent seeking, see Dennis C. Mueller, *Public Choice* II 229–44 (1989).

would function as a zero-profit monopolist. Such a contest, however, leads to an unproductive use of resources.

Theorists have long argued that efforts to acquire a monopoly waste resources. In fact, what we typically observe in the acquisition process seems to be a transfer of resources to others. Given that most if not all of the ultimate recipients of this transfer are themselves involved in legitimate activities, one might appropriately ask whether a real waste exists. There is a straightforward answer. Even if we assume that all of the resources devoted to the acquisition of monopoly status are transferred to people involved in legitimate activities (e.g., public-interested regulators, lawyers, economists), it remains true that the transaction ultimately ends in a transfer of wealth. Because that transfer adds nothing to the stock of goods and services, it seems appropriate to refer to the resources employed in effecting this transfer as wasted. In particular, activities that serve no purpose other than to create and maintain the monopoly position are thorough sources of waste.[16]

4. Inadequacies of the Criticism of Monopoly. Here I want to point out some rather sturdy criticisms of the antimonopoly position. Start with the short run–long run distinction. Joseph Schumpeter noted that short run profits must appear, for otherwise the incentive to enter and the incentive to innovate would never exist.[17] But the attainment of a monopoly position often provides short run profits. Take, for example, the process of innovation. A firm develops a new production process that results in a dramatic reduction in costs, leading to an expansion of output and a short run monopoly. Over time, others learn how to mimic the process and entry reduces profit to zero. Schumpeter's important insight was that the innovation-profit-entry process is a chain that is linked at several places, and that the causation is not necessarily unidirectional.

[16] There is a second and less clear answer: to the extent that beneficial externalities result from the wealth-transfer process, perhaps not all of the resources employed are wasted. Suppose, for example, that lawyers employed in the transfer process gain skills and develop law that may be usefully applied in other, productive sectors of the economy. In this case it is less clear that the entire process is wasteful. Still, it is probably a waste. To make the case that the transfer process is productive under these conditions, one must show that the resources used in the transfer process could not be used more productively in some other sector of the economy in which the same or similar beneficial externalities are possible. This is a high burden of proof. And if it is not, this merely serves as an example of how difficult it is, when one introduces externality considerations, to reach any firm theoretical conclusions on the desirability of monopoly.

[17] Joseph A. Schumpeter, *The Theory of Economic Development* 128–56 (1934).

The prospect of earning large, temporary profits generates efforts to innovate. Successful innovation leads to large profits. Take away or reduce the size of profits, and you will see less innovation, and less entry.

Schumpeter's point can be made somewhat clearer by considering the expectations of actors. Schumpeter's claim is that equilibrium requires an expectation of profit resulting from innovation. If innovation does not lead to profit, firms would not devote effort to innovation, because of its cost. However, if firms devote no effort toward innovation, the likely profits from innovating would be large. The prospect of short run profits must therefore always be present in a competitive economy. A second implication is that policies that tend to reduce short run profits also reduce innovation incentives.

The Schumpeterian argument remains a very strong one that receives too little attention from antitrust policy makers. Operationally, it suggests that we should be careful about enforcing the anti-monopoly provisions of the Sherman Act. Aggressive efforts to dissolve businesses with large market shares and high profits as soon as they appear may lead in the long run to a reduction in society's wealth.

A second argument against zealous enforcement is a variation of the preceding one. In certain areas, innovation provides spillover benefits to other firms and other industries. The innovator cannot collect compensation for the spillover benefits. Hence, the attainment of a short run monopoly is the best an innovator can do. Our patent laws already embody this theory.

Not every process innovation or neat idea can gain patent or copyright protection. Copyright protects expression, not ideas. Patent protection requires a certain degree of nonobviousness, and there are vast areas that cannot receive such protection, such as graphic designs with functional features, and mathematical formulae. In light of these large gaps in government protection, some incentive must be provided to innovators who fall within them. The prospect of short run monopoly provides this incentive.

The third argument is that the prospect of attaining a short run monopoly may spur entrepreneurs to seek out and identify consumer tastes not sufficiently satisfied by the range of products already on the market. The result is the introduction of new, differentiated products. The differentiated-product monopolist enjoys a short term monopoly. As time passes, others will enter the field until economic profits fall to zero. However, because these firms do not face an infinitely elastic demand