
Diagnosing monopoly (1979)

It is difficult being an expert in a subject everybody understands. It may even be more difficult to be an expert in a subject which everybody thinks they understand. In this regard, I do not envy football coaches, but being an economist has something of the same aspect.

Perhaps because economics deals with matters which touch everyone's life fairly closely, people tend to suppose that economic analysis must be just a matter of common sense requiring no special expertise. The difficulty is compounded by the fact that economists often use words which are in common use and whose everyday meanings are not in fact the same as their technical definitions. (This may be the only really good excuse that one can think of for the tendency of economists to create their own jargon.)

This sort of difficulty is perhaps most familiar to macroeconomists. The problem of explaining why a balanced budget does not mean the same thing for the economy as a whole as it does for an individual household is perhaps the best-known example. Yet the difficulty exists at the level of microeconomics as well. In particular, it is true of the problem which is the subject of this paper – the problem of deciding when a particular firm is or is not a monopoly. That problem, as indeed the word 'monopoly' itself, is surrounded by a miasma of not always consistent connotations. The man in the street, the legal profession, and the economist (not to mention competitors of the alleged monopolist) all have something different in mind when they talk of 'monopoly'.

4 *Antitrust: Monopoly and oligopoly*

Those different definitions need not always conflict, but they certainly tend to confuse. Moreover, the problem of deciding whether a real-life firm has monopoly power is further exacerbated by the fact that some economists take a very simplistic view of what it is that economic analysis has to say about the matter.

The world would be much simpler if one could determine that someone had a monopoly by observing that he owned Park Place and Boardwalk and was building hotels apace. (In that case, one might add, the monopolist could be sent directly to jail.) Similarly, the world would be a simpler place if one could decide that some firm has a monopoly by observing that it has 100 per cent of whatever it is that the firm is selling. After all, competition must mean that everybody who wants to can get in and survive, and monopoly, as everybody knows, means that one firm has it all. Unfortunately, that too is too easy.

In order to understand why such simplistic notions are wrong and to lay some foundation for examining the issues which arise, I shall begin with a brief review of the economic theory of competition and monopoly. Have done so, I shall then go on to consider the various issues which tend to arise in real-life cases. Generally, these can be grouped under the following headings:

- market definition
- the role of market share
- profits and their meaning
- barriers to entry and
- conduct and predation.

In almost every one of these issues we shall see that there is a certain amount of confusion in part engendered by the vocabulary and in part, perhaps, engendered by badly understood analytics. A little economics is a dangerous thing. What is true about a lot of economics, I am not prepared to say.

The pure theory of competition and monopoly

The static case

Imagine a market where there are many firms each selling the same homogeneous product for which there are no close substitutes, with each firm sufficiently small that it correctly believes that it cannot have an appreciable effect on the price. Suppose further that the consumers of the product are informed about prices and product quality, so that they know what is going on. Further, suppose that the owners of the

factors of production (the workers, the suppliers of capital equipment, and so forth) all understand perfectly well the contribution that they make to the production processes of the firms and can change firms with no difficulty.

In such an ideal world, any given firm will have no choice as to the price it charges its customers or pays its suppliers and workers. Seeking to maximize profit, it will produce at a point where marginal cost equals price – that is, at a point at which the price consumers are just willing to pay for another unit of the product exactly covers the cost of making it. Further, each supplier of a factor of production will be paid their marginal contribution to revenues, so that the firm, at the margin anyway, acts as a conduit passing what consumers are just willing to pay for another unit of the product back to the factors of production which just require that payment to produce it.

Will such a firm make profit? Well, that depends on what you mean by ‘profit’. The firm will certainly receive sufficient money to keep it in the business. Economists like to count that sort of thing as a cost rather than a profit. It represents, in effect, the return to the firm for its own services as a factor of production – the factor which puts all the other factors together and takes risks, so to speak.

Will profits be earned beyond this? Not in the long run. If firms are earning profits in the short run, then other firms, perceiving the opportunity to do likewise, will enter the business. That entry will expand supply of the product and bid down the price. It will also expand demand for the factors of production and bid up their prices. The end of the process will be a situation in which no profits (beyond the ones conventionally included in costs) are being earned.

What happens to inefficient firms in this situation? They cannot survive. Firms who do not adopt the most efficient means of production and are more costly than necessary, will find that they lose money. They are undercut by others whose costs are low enough to provide lower prices and they must either become more efficient or go out of business.

This suggests, and indeed it is true, that there is something good about competition. Not only is it true that firms must use the most efficient means of production, but also the marginal conditions already referred to ensure that consumers get what they are just willing to pay for, where the cost is computed using the most efficient means of production possible. Where such a situation prevails in all markets, it can be shown that it is not possible to rearrange things to make some consumer better off without making other consumers worse off.

Now let me complicate this simplified picture somewhat by introducing product differentiation. Suppose that there are different varieties of the same product, all close substitutes. Now firms must not only choose

how much to produce and how to produce it, but also what variety of the product to produce. Nevertheless, despite this complication, in broad outline the result will be roughly the same. Firms will produce any variety for which there is a demand creating a profitable opportunity. The end result of the process will be that consumers (who this time care not only about price but also about the type of product) receive a menu of product varieties each produced in the most efficient manner, each with the marginal cost of production just equal to its price, and each varying from the other in the sense that the cost of converting production from one variety to another just reflects at the margin what consumers would be willing to pay to make that change. (What has been left out of this picture, of course, is the process of invention of new product varieties and the question of how consumers come to learn about them. I shall take up those matters later.)

Let us contrast this happy state with that of monopoly. Suppose that we have a single firm producing the original homogeneous product for which there are no close substitutes. Such a firm would be foolish not to notice that the price it can charge depends on the amount that it wants to sell. The firm will not produce where marginal cost equals price; rather it will produce where marginal cost equals marginal revenue. This is because it will find it profitable to increase output up to the point where the cost of the last unit just equals the return from selling that unit, but unlike the case of the competitive firm, that return must take into account not only the price at which the unit is sold but also the fact that the price on all previous units will have to be lowered in order to sell all the units including the last one.

Another way of saying this is that although there are consumers who would be willing to pay the direct cost of producing another unit, they do not get the opportunity to do so, because the monopolist perceives costs as including not only the direct costs of manufacture but also the reduction in revenue on all previous units consequent on having to move the price down. Output ends up less and price ends up greater than in competitive equilibrium.

Does the monopolist earn a profit? In general yes. (I ignore the possibility that monopoly profits might be non-positive.) Unlike the competitive case, however, the profits can persist because entry by other firms does not take place. Without something blocking entry, no firm can be a monopolist.

Can a monopolist be inefficient? The answer is yes. If a monopolist is inefficient, less profit is made, but unlike the competitive case, making less profit need not mean making losses. It has been said 'the best of all monopoly profits is the quiet life' [Hicks, 1935, p. 8]. And a monopolist may in fact choose to take out the profits, as it were, in not aggressively

pursuing efficiency. Moreover, as the discussion of the marginal conditions indicate, the situation is not efficient in a larger sense. Even if the monopolist is using the most efficient methods of production, it would remain true that consumers willing to pay the costs of producing an extra unit will not be able to do so.

This simplistic contrasting of competition and monopoly forms a background for the rest of this paper. It should be remembered in proceeding that it is simplistic and that my purpose is to concentrate on the diagnosis of monopoly. There are, in fact, a number of intermediate market structures, such as oligopoly, which present their own problems. My remarks should not be taken as applying to them. This is perhaps particularly so when introducing the first complication of the simple model, the consideration of innovation and change.

Innovation and change

So far, this discussion has been almost exclusively static, essentially comparing points of equilibrium. But competition (or the lack of it) is a dynamic process and takes place in time. Since we typically only obtain snapshots of a moving process rather than observing it only after equilibrium has been reached, it is important to consider the way in which things change over time. This is particularly true in industries characterized by large amounts of innovation.

What does competition in such an innovative industry look like? For simplicity, I concentrate on innovation which takes the form of the introduction of new goods, although much the same story could be told of other types of innovation and, in particular, of the discovery of more efficient methods of production.

Start from a position of long-run equilibrium in a competitive industry. Revenues balance costs; no profits are being earned. Now someone – the innovator – discovers a better product and brings it to market. If the product is truly better (indeed, this is a definition of what ‘better’ means), consumers will prefer it to the existing products. They will be willing to pay more for it. This means that, during the initial period as the only one producing this product, the innovator will be able to charge a high price for it and will earn profits.

Just as in the static model, however, the presence of such profits will lure others into the business. If there are no barriers to imitation, other firms, through reverse engineering, for example, will learn how to make the new product and begin bringing it out. This will bid down the price, since some price advantage will generally be necessary to lure customers from the original to the copy. Moreover, the imitators will generally

have lower research and development (R & D) costs than did the initial innovator and they will be able to afford to get into the imitation business in the first place at a price lower than that which would have brought a reasonable return for heavier R & D expenses. At the same time, the general progress of technology or simply the experience of those who have gone before may enable imitators to make slight improvements on the product and other innovators will make further improvements, bringing in newer and better products. This too will cause customers to leave the original innovator.

Faced with the erosion of business and profits caused by the entry of imitators and rival innovators, the original innovator will not be able to maintain the price which brought the profits of the initial period. In order to stay in business, the price must be lowered on what is now the old innovation and, to make still further profits, still better products must be brought out.

In some respects, the story just told is not very different from the static competitive case. Once again, the lure of profit induces entry and that entry bids the price down and the profits away. Once again, firms are forced to seek the most efficient means of production or, in this case, the best product, or leave the business. Yet the innovative competition story clearly highlights somewhat different points than does the static one and they are worth taking a minute to point out.

The first of these is the necessary role of profits and what they represent. Profits in the initial period of the competitive process just discussed are the reward to the innovator. They are what caused investment in innovative activities to be made and they represent a return on those activities. The US Constitution recognizes the importance of such returns in encouraging innovation in the form of the patent system.

Related to this is the fact that one cannot look at profits during the initial period and attribute them only to the manufacturing process (this is in addition to all the problems associated with using accounting data for profits which I shall discuss later). The profits being earned represent not merely the return on capital invested in manufacturing but also on the investment made in research and development.

Third, and perhaps most important of all, it is a mistake to look at an industry in the midst of such a process and conclude anything at all about it without considering where the process came from and where it is going. A single frame will give a misleading impression of the movie. Looking at the industry during the period just after the innovation is made, one sees the world beating a path to the door of the mousetrap inventor. One sees the mousetrap inventor making profits. One sees the mousetrap inventor alone in the field. One ought not, however, to

conclude therefore that a monopoly of mousetraps exists. Indeed, what really matters, in some sense, is whether there is a monopoly of technical progress in the industry. Similarly, when prices come down after the imitators enter, it would be wrong to conclude that the monopolist is engaging in 'predatory pricing' in order to maintain market share. Rather, what one is seeing is competition seriously at work.

What would a monopolist look like in such an industry? In effect, a monopolist would be someone, as in the static case, insulated from entry, from the pressures of imitators and other innovators. New products might very well be brought out because that would be a profitable thing to do, but the crucial fact would be that it is not necessary. As in the static case, the crucial difference between monopoly and competition is the compulsion which market forces place on the competitor and the lack of it on the monopolist.

Monopoly and monopoly power

Since the case of a monopolist with 100 per cent of an economically relevant market (see what follows) is very rare, the analysis of monopoly in the economics of antitrust policy tends to go in terms of 'monopoly power'. This is entirely proper. Such power is in fact the lack of compulsion we have spoken of. The courts have defined monopoly power as 'the power to set prices and exclude competitors', and from the point of view of economic analysis, that is an excellent definition if it is properly understood. Clearly, a monopolist has the ability to earn profits while excluding competitors. This generally means setting high prices while excluding competitors. Whether one ought to infer monopoly power from the ability to *cut* prices and thus exclude competitors is another matter to which I shall return.

The question of identifying a firm with such monopoly power when observed in the field, however, is not a simple one. I turn now to a deeper discussion of some of the problems which beset the enthusiastic monopoly-watcher.

Market definition

The conventional first step in analyzing whether a given firm does or does not have monopoly power is to define the relevant market in which the alleged power is exercised. Unfortunately, this is not as simple as it sounds and tends to lead to confusion, if not abuse.

In my discussion of competition, and especially monopoly, I simply started off by assuming that there was a single homogeneous product with no close substitutes. Now it is all very well, if that happens to be the case in real life, but one frequently encounters the problem that products are differentiated. Typically, in such cases, we are talking not about a single product, but about a group of products. Market definition might be described as the problem of deciding where the relevant group begins and ends. This turns out not to be an easy task and I shall point out in a moment that it may be an unnecessary one, yet common sense (and Supreme Court decisions) suggest it has to be undertaken. After all, one might say, in order to decide whether a firm has monopolized, it is necessary to decide *what* it may have monopolized. In order to decide whether a firm can exclude competitors, it is necessary to decide from what they can be excluded.

The position which I shall take here is that, properly done, trying to answer such questions does indeed yield information about monopoly power. The catch lies in the words 'properly done'.

Let us begin by recalling what the purpose is of market definition. It is the beginning of an analysis of monopoly power. Monopoly power, however, is the ability to act in an unconstrained way. Hence, market definition, if it is to be an aid to analysis, has to place in the relevant market those products and services and firms whose presence and actions can serve as a constraint on the policies of the alleged monopolist. Recall that market definition will be used essentially for two things (both of which I shall discuss later): the computation of market share and the analysis of barriers to entry. A market will thus be well defined if and only if the share measurements to which it leads provide some reasonable index of the true power of the alleged firm; the discussion of entry really then supplements the share measure to show the ability of that firm to maintain share when earning supranormal profits.

Thus, the primary question in defining a relevant market ought to be that of the constraints on the alleged monopolist. The principal constraints can be of two types, those relating to demand and those relating to supply. The courts have paid appropriate attention to demand and supply substitutability – appropriate because those are criteria by which to judge the constraints on the alleged monopolist. It should not be forgotten, however, that it is the constraints which are the object of analysis and not the properties of substitutability themselves.

Demand substitutability refers to the ease with which consumers of the alleged monopolist's products can substitute the products of others. If this is relatively easy, then an attempt by the alleged monopolist to raise prices and earn supranormal profits will lead consumers to switch away. Notice, however, that such substitutability is a question which

depends on relative prices. Any firm which raises its price far enough will lose customers. The issue is whether over a range of actually encountered prices consumers are in fact able to substitute.

Note further that such substitution is often not as simple as meets the eye. Consider the following example. My wife and older daughter who (like the rest of the Fishers) enjoy skiing, have chosen to use for their bindings (the device which holds the ski boot to the ski and releases in an otherwise dangerous fall) the Spademan binding which has a particular design. It happens that, because of the design of the Spademan, there are relatively few ski boots which are compatible with it. Many otherwise desirable boots cannot be used with the Spademan binding. Would it be correct to assume that Spademan-compatible boots constitute a market because owners of Spademan bindings cannot directly substitute non-Spademan-compatible boots for Spademan-compatible ones?

The answer is that it would not, even restricting our attention to demand substitutability. It is true that once my wife and daughter acquired the Spademan binding, they limited their immediate range of substitution. An attempt by the makers of Spademan-compatible boots to raise prices and earn supranormal profits would not and could not lead my family to try to ski with Spademan bindings and non-Spademan-compatible boots. But that kind of substitution is not the only kind which constrains the behavior in question.

In the first place, a sufficiently high price for Spademan-compatible boots will lead my wife and daughter to discard their Spademan bindings and acquire other bindings. Moreover, such a price need not be terribly high – the price of the boot is often more than 50 per cent of the boot-binding combination. Second, below that is a price at which, when their current bindings wear out, they will replace them with non-Spademan bindings rather than have to pay the high price for Spademan-compatible boots. Third, and more important, new customers, deciding on what binding-boot combinations to buy will rationally look at the high price of Spademan-compatible boots and make the binding-boot choice taking that into account. It is a great mistake to look at substitution as though all that mattered was rolling out one product and rolling in the other once everything else is fixed. Rather, a major kind of substitution occurs at the stage before everything gets fixed. To the extent that new (or replacement) binding customers are important, the makers of Spademan-compatible boots will have to think twice before attempting to take advantage of those who have temporarily acquired Spademan bindings.

The same example may be used to illustrate some other principles. To the extent that certain boots are associated with certain bindings (I

am wandering somewhat from what is really true about ski boots), the real competition takes place between binding–boot combinations. It would be wrong to consider the market for boots alone, even if boots are sold without bindings, if there is a substantial business in binding–boot combinations and the price of the boot affects the choice of the combination.

But what if there are some people who simply want a particular boot because they feel committed to the use of a particular binding, for example? One might add, what if there are people who particularly want a chartreuse boot with seven buckles and a monogram? Does that mean that such people constitute a relevant market and that the maker of such a boot (if there is only one) has monopoly power? It would be a mistake to conclude this automatically. Once we leave the theoretical world of perfect competition and examine real-life firms with differentiated products, we find that every such firm tends to have customers who would be willing to pay more for the product than other customers. This is what it means to have a downward-sloping demand curve. Yet it would be foolish to conclude that this means that there are lots of little markets, each one consisting of such customers, with the corresponding firm having monopoly power. There is no relevant sense in which that is true. To content oneself with the question of whether there are any substitutes for a given product which can be used by those who happen to want the particular product, with exactly the specifications and properties that that product has, is to ignore the real forces at work in the market and to beg the question of market definition which arises when products are differentiated.

The second kind of consideration in market definition is that of substitutability in supply. Let us continue with the ski boot example. Suppose that the makers of non-Spademan-compatible boots could very easily produce Spademan-compatible boots if it were profitable to do so. In that case, an attempt by the makers of Spademan-compatible boots to earn supranormal profits would induce other bootmakers to change their production and bid away those profits. In such a case, it is not very sensible (even apart from demand substitutability considerations) to talk of a market for Spademan-compatible boots. To do so and then to look at the market share as indicating anything, is simply to ignore the important constraints on any power of Spademan-compatible-boot manufacturers placed by the presence of the other boot manufacturers. It would obviously be more sensible to count the latter in the same market.

Here again, it is important to realize what kind of substitution is involved. The issue is not whether, once a non-Spademan-compatible boot is produced, the manufacturer can easily convert it into a Spade-

man-compatible one by retrofitting. If that is possible, of course, it is important. Rather, the issue is whether the production facilities can be adjusted to make Spademan-compatible boots.

Obviously, such substitutability is a matter of degree. Where supply substitutability is somewhat more difficult or slower, we may prefer to draw the boundary of the relevant market and refer to the ability of firms outside it to make the product inside it in terms of ease or difficulty of entry, a matter I shall take up later. Indeed, properly done, it matters not at all where we draw the market boundary from the supply substitutability point of view. If we draw it very narrowly, we shall have to say that entry is extremely easy. If we define the market much more broadly, we shall have to remain aware that not everyone in it puts an equal constraint on the power of the alleged monopolist.

As this last point suggests, I do not believe that the question of what is the relevant market is the fundamentally right question to ask, even though answering it in a sensible way can be an aid to analysis. The fundamental question is that of the constraints on power. Focusing on the question of relevant market can often lead to losing sight of that fact.

Let me elaborate. The inevitable next step which comes after a market has been defined is the computation of market share, about which I shall have more to say in the next section. Ever since Judge Hand's dictum in the *Alcoa* case as to the various market shares which might lead one to infer or not infer a monopoly, plaintiffs have struggled to prove shares to be higher and defendants have struggled to prove them lower than the points he named. Obviously, if a market is defined sufficiently narrowly, one can do it so that the share of the alleged monopolist is high. Similarly, if the market is defined sufficiently broadly, one can get the share of an alleged monopolist to be lower.

Yet all this is in some sense beside the point. Whatever shares may mean, their meaning depends on how the market is defined. In a market defined overly narrowly, a high share does not carry much information. Similarly, in a market defined overly broadly, a low share also does not. In the former case, not all constraints on the alleged monopolist's behavior have been taken into account. In the latter case, not all the things which have been taken into account constrain that behavior equally.

If one always remembers this, there is no positive harm in engaging in the market definition exercise. Indeed, viewed correctly, arguments concerning whether products are in or out of the market which are made in terms of demand and supply substitutability, and hence in terms of constraints, are exactly the arguments which one would have to decide in looking at the constraints directly. The trouble is that it is too easy to

forget what the analysis is all about. By focusing on whether products are in or out of the market, one converts a necessarily continuous question into a question of yes or no. The temptation is to regard products which are in as all counting equally and products which are out as not counting at all.

The result of this tendency in antitrust cases is for plaintiffs to argue for narrow market definitions not in terms of constraints and their implications but in other terms altogether. Thus, where concentration should be on the competition faced by the defendant, private plaintiffs who compete with the defendant in some part of the latter's business try to define the market in terms of the competition they themselves face, something which would tell one about the constraints on the plaintiff but not about the constraints on the defendant.

Moreover, as the foregoing discussion should indicate, the term 'market' as used in this kind of analysis is a term of art. It is not the same as the ordinary use of the word. Yet there is a tendency to adopt ordinary usage as though it had technical meaning so that a tendency of business executives to speak of the 'market' for a particular kind of their own product becomes converted into a supposed recognition that such is the relevant market for economic analysis. This is not only bad practice, but it would make economic experts unnecessary and is therefore to be frowned upon.

Such a practice is closely related to another one. When dealing with records or statistics on differentiated products, it is natural to categorize those products and to treat each variety separately. Yet such record-keeping treatment cannot convert the categories involved into economically relevant markets for the analysis of monopoly. Similarly, one cannot look at the definition of words and construct markets. To the extent that market definition really is a matter of definition, it has to be only a way of looking at the problem – a way which cannot affect the answer. To the extent that it reflects a real part of the analysis, one must bear that always in mind.

Let me take an example from a real case about which I know almost nothing (as it happens, this is a merger, rather than a monopoly case, but the principle is the same). It is my understanding that the Federal Trade Commission has challenged Nestlé's acquisition of Stouffer's and has claimed that the relevant market is that for high-priced, non-ethnic, frozen entrées (main courses). Let me just take a moment to be sure that that sinks in. High-priced, non-ethnic, frozen entrées. Not only are all non-frozen foods excluded, but so are any frozen dinners (or combination of frozen entrée and unfrozen vegetables), anything such as meatpie (low-priced), and any Chinese, Mexican or Italian food (ethnic). I have no opinion whatsoever on whether or not the merger

involved tends to reduce competition, but it is not necessary to know that in order to see that if adults wish to make the question of whether high-priced, non-ethnic, frozen entrées is a relevant market, the question on which the analysis turns, then somebody's eye is badly off the ball. One suspects that whoever made that up has lost track of what the whole business is supposed to be about. (It is amusing that sometime after the case was brought, Stouffer's ran a series of commercials on television which showed someone tasting a Stouffer's entrée and saying things such as 'It's like lasagna, but it isn't lasagna'. Whatever the motivation behind such commercials, the implication is that certain ethnic entrées compete directly with Stouffer's products.)

Market share

As already observed, the big point in defining a relevant market is to proceed to the computation of market share on the supposition that this tells us something about monopoly. After all, everybody knows that a pure monopolist has 100 per cent of the market. Presumably the higher the share the more likely the inference of monopoly power. Is this correct?

I have already emphasized the importance of getting the market definition part of the analysis right if one is going to try to make inferences from share. The computation of share in a market with differentiated products is an attempt to summarize a complex set of relationships and to read into a single number a number of items of somewhat different weight. If market definition is done properly, that can be an aid to analysis, so long as one remembers that one is summarizing and possibly leaving out important information.

Suppose then that market definition has been properly done (whatever that may mean). What does economic analysis tell us about the relation of share and monopoly power? Well, the one proposition which most people believe is that a small share shows the absence of monopoly power and a large share its presence. (Note that I am careful not to say how small is small, however.) This is not true. The right question is that of what happens to share, or, more generally, to a firm's business when monopoly profits are sought. The fundamental issue is whether competitors are able to grow.

Thus, consider a firm which has a very large share of a particularly defined market. It may very well be that such a firm is merely efficient and has achieved that share by charging low prices. Alternatively, we may be looking at a case of innovative competition in the initial period when the mousetrap has been invented. Should we infer monopoly

power from a large share in such cases? The answer is no, not necessarily. The right question to ask is whether that large share would survive an attempt to charge high prices and earn monopoly profits. If the share is maintained solely because of low prices or better products, then we are looking at what competition is supposed to do and not at a monopoly. This is, of course, closely related to the legal position that a monopoly acquired and maintained by 'superior skill, efficiency or foresight' does not violate the antitrust laws. I would prefer to say that a large share acquired and maintained in such ways is not a monopoly at all.

The confusion of monopoly with large share is dangerous in complicated cases. When combined with the related concentration on market definition as the great touchstone question, it leads to analytic confusion.

Profits

If one cannot be sure one is observing a monopoly by observing a large share, are there any other simple features of monopoly that might enable one to conclude that it is present? The one most commonly pointed to is profits. My discussion of share, just given, suggests that the crucial issue regarding that variable is what happens to it if supranormal profits are earned. The discussion of the foregoing competitive case shows that it is the lure of profits which leads to entry which then bids the profits away, a matter I shall take up later. Is it then true that one can look at a firm's profit rate and conclude very simply that there is monopoly power if it is high and an absence of such power if it is low?

The answer is no, and I am sorry to say that there are a number of studies in the literature which attempt to examine the relation between profits and concentration which automatically equate profits with monopoly and tend to make other errors as well. Such issues also creep into antitrust cases in analytically foolish ways.

The analytically correct issue here is the measurement of what is called 'the economic rate of return' and its comparison with some competitive standard. The economic rate of return is a profit rate relating profits to capital (profits on sales are not involved). It is defined as follows.

Consider an investment which costs a certain amount and which over time brings in a stream of net revenues as a result of having been made. (Net revenues are the difference between gross revenues and the costs associated with maintaining the revenue stream.) The economic rate of return on the investment is that interest rate such that, using that rate, the discounted present value of the stream of net revenues is just equal to the capital cost of the investment. It is, in effect, the interest rate

such that, if one could get it at a bank, one could deposit the capital value of the investment and just manage to draw out the same stream of net revenues as the investment generates. It is the expectation of a high economic rate of return relative to similar opportunities elsewhere which draws capital into a competitive market.

Obviously, calculating the economic rate of return in any but the simplest situation is not a simple matter. Calculating it for a firm as a whole involves knowing a great deal about the past, present and, indeed, future history of the firm. Yet the important fact for our purposes is that calculations involving so-called 'accounting rates of return' which can be read fairly easily from the firm's books are no substitute whatsoever.

This is true for a number of reasons, some obvious and some more subtle. The accounting rate of return is essentially current profits divided by some measure of the value of current assets. (Sometimes it is taken as current profits divided by the value of stockholders equity. I ignore this version, for simplicity.) The most obvious reason that such measures can be misleading concerns the question of what is included in profits and what in the value of assets for accounting purposes. This involves such questions as the treatment of depreciation for accounting purposes as opposed to true economic depreciation, but it involves other issues as well.

To take one of the easier issues, many firms choose to write off research and development or advertising expenses as part of current costs. This is a conservative method of accounting and may in fact be useful for tax purposes. Yet research and development expenses or even advertising may be analytically equivalent to capital expenses. I do not know what du Pont's accounting practices were, but it ought to be plain that the very large research and development expenses for the introduction of nylon, for example, carried on over several years, were analytically equivalent to an investment producing a stream of much later returns. Properly calculated for the purpose of assessing the economic rate of return, such expenses should be capitalized and placed in the asset base rather than subtracted from current revenues. The effect of making such a change on the calculated rate of return is complicated.

There is a much more important problem with the accounting rate of return of which, in some respects, the one just discussed is a subcase. It is easiest to highlight this with a numerical example. Suppose that a typical investment consists of the purchase of a machine costing \$100. Suppose further that such an investment brings in nothing the year it is made but, starting one year later, brings in \$11 per year in perpetuity. It is not hard to see that the economic rate of return on such an investment

has to be 10 per cent. (That is, \$110 invested at 10 per cent will bring in \$11 per year in perpetuity. Thus, the stream of returns is equivalent to \$110 invested *one year after the machine is bought* at 10 per cent. But \$110 one year away is the equivalent of \$100 now at 10 per cent so the rate of return on the whole thing must be 10 per cent.)

Now, I have chosen 10 per cent to make the numbers easy, but let us suppose that 10 per cent is in fact a high enough return so that the firm wishes to make such investments and to make them at quite a substantial rate. Suppose that it does so as follows. In the first year, it buys one machine. In that year, capital stock is \$100 and profits are zero; the firm's accounting rate of return is thus also zero. In the second year, it buys another machine so that it doubles its capital stock. Since there is no depreciation in this example, the value of the capital stock is now \$200, but the firm begins to earn the \$11 per year brought in by the first machine. Its accounting rate of return is now 5.5 per cent. In the next year, suppose the firm again doubles its capital stock, investing in two more machines and bringing the total value of its capital stock to \$400. It now earns \$11 from the first two machines for a total of \$22 and an accounting rate of return still 5.5 per cent. In fact, if the firm goes on doubling in size every year, its accounting rate of return will always (except for the very first year) be 5.5 per cent, about half of the true economic rate of return.

Evidently, from this example, the relations between the accounting rate of return and the economic rate of return depend a great deal on the rate of growth of the asset base of the firm. At least in this example, the faster the firm grows, the lower the accounting rate of return will be, leading to the anomaly that very high economic rates of return which would induce the firm to grow very fast will produce very low accounting rates of return. Moreover, outside of such a simple example, the problem is not always one way. Depending on the rate at which firms invest, the rate at which investments depreciate and the time pattern of the returns from an investment, it is quite possible to have a low economic rate of return correspond to a very high accounting rate of return. One cannot make inferences about monopoly profits from this sort of thing. The correct analysis is not a simple one.

I regard the problem of calculating the economic rate of return as the most serious one in trying to use the profit rate for a judgment about monopoly power and the use of the accounting rate of return as the worst mistake therein. Nevertheless, even if one somehow succeeded in getting around this sort of timing problem in estimating the economic rate of return, there would still remain problems with using the results to make inferences about monopoly power. As with some of the problems raised previously, the issues here involve what is appropriately called profits.

Recall that, in the competitive model, a certain amount of what accountants and firms consider profits is included in costs for the purposes of economic analysis. Expressed as a rate of return, the amount so included represents the rate of return just necessary to keep the firm in the business; it represents the rate of return that the firm's capital could earn elsewhere. Thus, when one says that long-run profits in a competitive industry are zero, one does not mean that measured profits will be zero, but merely that measured profits will be at that level at which there is no inducement to entry or exit.

Now that is all very well in theory, but in practice, one is going to make a judgment about monopoly power by deciding whether the given level of profits is at a competitive level or whether it is above that level with entry somehow blocked. This would be easier to do if there were a single measure of a competitive rate of return.

There is no such single measure, however, and this is largely due to the fact that risks in different industries vary. When I outlined the competitive model, I took the usual course of ignoring risk altogether. But real-world industries, even real-world competitive industries, experience risk and part of the rate of return necessary to keep capital in the business is a reward for risk-taking. Accordingly, the competitive rate of return in high risk industries is going to be higher than in low risk ones.

Nobody knows exactly how to quantify that effect, yet it is obviously an important issue. Consider, for example, a very high risk industry in which those who gamble and fail leave. The profit rates earned by those who gamble and succeed and thus remain in the industry will be high, yet it would be wrong to conclude that one is looking at a situation of monopoly profit. If one looks only at the rate of return earned on their investment by winners of the Massachusetts State Lottery, for example, one will quickly conclude that it is an extremely good investment (a view which the advertisements put on by the state would like to encourage). Yet this is clearly wrong. What matters is the rate of return for both winners and losers.

Moreover, there is an additional, somewhat related, issue. Suppose that one looks at an industry and manages to calculate the economic rates of return for the firms therein and finds that there is one firm with a very high rate of return and others with rather lower ones. Can one then conclude that the most profitable firm has monopoly power? After all, in competition, are not all firms' profits supposed to be the same and equal to the competitive rate? Even if one cannot tell what that rate is, cannot one conclude something from differential profits?

Unfortunately, the answer is once again in the negative. In the risk-taking case, this is because there may be differential outcomes for winners and the rate of return earned by the luckiest winner may be more than that necessary to remain in the business. Even so, monopoly

profits are not earned because what is involved is the *ex ante* expected rate of return involved when the gamble is taken. More generally, what is involved here is what economists call 'unimputed rents'.

The simple model of competition supposes that every firm is just as efficient as every other firm and has access to all the same factors of production. Yet in practice, this may not be the case. One firm may be more efficient than another. To take three examples, it may have more efficient people working for it; it may be more favorably situated geographically; or it may at some time in the past have made a successful innovation which, perhaps just because it was first, cannot be successfully duplicated.

How does one reconcile such cases with the competitive model? One way to do it is to define the apparent excess profits being earned as really being something else. This is the sort of thing that economists are very good at, and so it is in this case. Such apparent profits are called 'rents'. They represent returns which do not affect economic decisions and which can be thought of as belonging to something other than the firm's manufacturing activity. Thus, if the firm's management is particularly skillful, the extra money coming in really represents returns to managerial talent rather than profits to the firm. Similarly, a firm with an advantageous location ought to be thought of as making its extra money as a return on that location – a true 'rent' which the firm pays to itself as a landlord. Finally, the firm with the past successful innovation is earning money as the owner of the rights to that innovation rather than in its current production activity.

Now, in some circumstances, one would not expect to see such rents appearing in the profit statements of firms. If greater efficiency is due to the skills of only a few people, then, in a competitive industry, such people, if they are not paid for their skills, can be bid away by new entrants or other competitors. The end result will be that their wages go up and apparent profits go down. Similarly, in some very long-run sense, a firm with an especially advantageous location could lease that location to another firm at a high price and go into business at a less advantageous location. If it did that, it would record the same amount of incoming money, but what appeared to be profits would be properly recorded as rent coming into the firm in its landholding capacity. Analytically considered, the firm ought to keep its books in that form whether or not it actually engages in such activities. Similarly, the holder of the rights to an innovation ought to value those rights at what they can be sold for and keep books in two capacities, one as a manufacturing firm and one as the holder of such rights.

But, of course, firms do not keep their books in such ways. Moreover, even the case of special managerial skills need not result in

rents being fully imputed to the factors of production with which they are properly associated. Particularly in large firms dealing with complicated and delicate technologies, it is perfectly possible for the added efficiency to accrue not to any small group of individuals but to the firm as a whole. If that is true, then while it would be possible for others to bid away any small group of individuals, the managerial efficiencies would still rest in the organization, the whole being greater than the sum of its parts. In that circumstance, there would still be unimputed rents which will show up as profits in the accounting records.

Obviously, I believe that judgments about profits as an index of monopoly power are very difficult, if not impossible to make. This is particularly unfortunate because of the associated difficulty it produces in assessing barriers to entry, a subject to which I now turn.

Barriers to entry

It should be clear from what has been said that a consideration of the role of entry plays a major part in any assessment of monopoly power. Where entry is easy, no monopoly power can persist. Where entry is difficult, provided there are not already many existing competitors, monopoly power can survive. It is entry which is induced by supra-normal profits in a competitive industry and entry which bids those profits away. (Actually, this is literally true only if one reads 'entry' to include the expansion of existing competitors.)

Clearly then, correct analysis of entry or barriers to entry lies at the heart of an assessment of monopoly power. This is particularly true, since, as we saw when discussing market definition, the choice of an arbitrary line for the boundary of the market on grounds of supply substitutability will not affect the outcome provided that one is careful about entry from entities left outside the line. Whether considered as a phenomenon of new firms coming into the business or a phenomenon of older firms able to expand (when the line is drawn rather more widely), the analysis of entry conditions is the analysis of a central phenomenon which places or does not place constraints on the behavior of the alleged monopolist.

It is therefore with some regret that I have to say that the analysis of barriers to entry is, in my view, the single most misunderstood topic in the analysis of competition and monopoly. Even the confusion associated with market definition probably takes a second place. In large part, this may be due to an unfortunate terminology, but whatever the cause, it is a matter for considerable concern.

To see what the problem is, it is important to understand the

economic relevance of a barrier to entry. A barrier to entry exists when entry would be socially beneficial but is somehow prevented. That is a fairly fancy way of describing a situation in which unnecessarily high profits are being earned and society would be better off if they were competed away, but firms cannot enter to do this. The social benefit–cost calculation is not correctly reflected in the private benefit–cost calculation of the potential entrant.

Now, that definition, which is drawn in terms of the results one wants entry to have, departs rather sharply from what one might think would be meant by a ‘barrier to entry’ in terms of something which makes entry difficult. Nevertheless, it turns out to be the useful way to look at it. (I have been aided greatly in thinking about this problem by the recently published work of C.C. von Weizsäcker, 1980.)

Consider, for example, an industry, entry into which requires the construction of a large plant, a distribution network and other large investments. Leave aside for the moment the question of whether the scale required for entry is big relative to demand and assume that firms can borrow at rates which correctly reflect perceptions of risk, both matters to which I shall return. The question I want to focus on now is that of whether the mere fact that a large amount of money has to be spent to get into the business and that certain skills and equipment have to be acquired means that there is a barrier to entry.

This is where economic analysis and ordinary language part company. It is not true that the situation I have just described is necessarily one of high barriers to entry. Just to focus attention, let me point out that there are not necessarily high barriers, even if the incumbent firms are currently making very high profits. Why should this be the case?

An economically relevant barrier to entry is one in which unnecessarily high profits are not bid away by entry. That is a situation in which society would be benefited by entry but in which the attractiveness of entry from the point of view of society is not the same as attractiveness from the point of view of the entrant. In the large investment example just described, however, there is no reason to believe that it is necessarily the case that profits, *properly considered*, are excessive. If a large lumpy investment has to be made to get into the industry, then the right consideration for an entrant is what the rate of return will be on all the expenses, including that lumpy investment; but if the incumbent firms had to make similar lumpy investments, then consideration of the rate of return being earned by those firms (even apart from many of the issues raised in the previous section) has to take into account the fact that their current apparently high profits come as the result of having made those investments. What has to be considered is the rate of return

being earned by the incumbent on the entire process, including the initial investment. To look only at the short-run profits after that investment has been amortized for tax purposes, for example, is to miss the point.

Let me put it another way. If it is technically necessary to make a lumpy investment to get into the industry, the right question from the point of view of society is whether or not the rate of profit to be earned on the entire activity of entry, production and sale, including that lumpy investment, is higher than the rate of return which can be earned in other industries (adjusting for risk). The fact that a large investment has to be made at the outset makes this a long-run question, but it is the right question nevertheless. In the situation described, there is no reason to suppose that the calculation made by the potential entrant is any different from the calculation which society would wish to be made. There is no reason to believe that there are any social benefits or costs which are not reflected in private incentives. In the situation described, the potential entrant will not enter if the profits foreseen in the long run will not be sufficient to justify the initial lumpy investment. But that is exactly the same calculation which one would make on behalf of society. It would be socially wasteful to encourage such investment if the resulting profits will not be sufficiently high to earn the rate of return which could be earned by investing the resources elsewhere.

The same general principle applies to other situations sometimes thought to represent barriers to entry. Take, for example, the situation in which an existing manufacturer has achieved a deserved reputation for reliability and quality of product. Suppose that a new manufacturer, even though able to produce a product which is technically as good, or even better, cannot induce customers to buy this alternative except by offering a price premium. Does that disadvantage signal a barrier to entry?

The answer is no. Customers, when they buy a product in such a circumstance, buy not only the physical characteristics of the product but also take risks as to its quality and its reliability. One way of looking at it is to say that in buying the established products, customers are buying less risk or, if you will, more information. Investment in providing that service has already been made by the incumbent. The new entrant can also provide it by offering a price premium which will induce customers to try the new product and thus gain a reputation. But it is not to society's advantage for customers to be forced to take risks with untried products in such a situation. Just as in the case of the building of a plant, the question is whether the profits to be earned by ultimately attaining a reputation equivalent to that of the incumbent are such as to justify the investment which will have to be made to attain it. If so, then

both society and the entrant will be served by entering; if not, then neither will be.

This is an important and easily misunderstood point, so let me distinguish it from some related matters. In the first place, I am talking about deserved reputation – reputation which correctly reflects product quality. It may also be that customers have irrational brand preferences. It is hard to describe exactly what that would mean, but it would certainly be different.

Second, there is no doubt that society would be better off and entry easier, in some sense, if information were perfect and the risks of trying different products either did not exist or could be assessed free of cost. The fact that experience is necessary to reduce risks certainly makes the incumbent better able to earn profits than the potential entrant. Further, competition would be enhanced if the information and risk problem did not exist. Given that it does exist, however, there is no true barrier to entry in the fact that the incumbent, but not the potential entrant, has already invested in overcoming it. Exactly the same thing is true of investment in plants. If it were not necessary to build a plant to get into a business, then more people would be in it. In a world with that kind of technology, competition would be different from what it actually is. However, what makes the difference is not the fact that incumbents have already invested in the plant necessary to overcome the difficulties. That is the cure, not the disease. Given that such an investment has to be made, either to build a plant or to establish a reputation, the question both for society and for potential entrants is whether it is worth making.

Now, the fact that I have been careful to point out that two things which are commonly thought to be barriers to entry are not really that at all does not mean that barriers to entry do not exist. There certainly can be such barriers. Here are some examples.

Incumbent firms may possess all there is to possess of some scarce resource. It may be that they are not even using that resource to capacity but are, as it were, stockpiling it. In such an event, society would be better served if entry could take place and expansion of output occur.

Next, although the need to make a large investment may not in itself be a barrier to entry, it may be that it is associated with something that is a barrier to entry. Suppose that the minimum scale which is necessary for efficient production is large relative to demand. In such a case, an incumbent firm may be able to earn monopoly profits because an entrant will properly make the calculation of what profits will be *after* entry, rather than before, and will perceive that with one more minimum scale firm in the market the addition to supply will be such as to reduce prices to a point where profits cannot be earned. Note that, even here, if

costs decrease sharply up to minimum size and the market is not big enough to support one more firm when all firms are of minimum size, then the social benefit–cost calculation is the same as the private benefit–cost calculation and one ought not to talk of barriers to entry. In somewhat less black and white situations, however, what matters in part is what the entrant believes will happen to incumbents' outputs upon entry. There can be cases in which society would be benefited by the expansion of output consequent on entry, but in which entry does not take place because the entrant believes there will not be enough room. (Conduct-related issues are discussed in the next section.)

It may also be the case that credit markets function imperfectly, so that the rate of interest which must be paid by a small firm to make the lumpy investment required for entry is higher than that which would be paid by an incumbent firm to make a similar investment and higher in a way which does not simply reflect the greater risks involved. In the limit, it may be that a potential entrant cannot borrow the money at all.

The question of the existence of such capital barriers to entry is a matter of some dispute. I shall only touch on some of the issues. Aside from the difficulty of assessing whether differential borrowing rates actually reflect differential risks or not (if they do, then society's interests are appropriately reflected in the rates), potential entrants into many industries are not small firms but large firms who operate outside those industries. Since credit rationing, to the extent that it exists, is generally supposed to involve the size of the borrower rather than the size of the borrower in a particular business, the importance of capital barriers in practice has been questioned.

Further, it is important not to make mistakes by looking at the extent to which incumbents can finance expansion with retained earnings while new entrants have to raise new equity or borrow. That is not in itself an issue of the imperfection of capital markets. An incumbent who uses retained earnings to finance expansion is forgoing the opportunity to invest those earnings elsewhere, thus paying what is technically known as the 'opportunity cost' of investing in the business. Retained earnings are not free. It may, of course, very well be that internal financing is cheaper than external financing. But this may simply reflect differential assessments of risk rather than a true capital barrier. It is not an easy thing to decide.

There may also be conduct-related barriers to entry; I take these up in the next section.

In general, in deciding on barriers to entry, it is important to look at the whole picture. In particular, one has to look at the long-run picture. Take the case of innovation. After the mousetrap has been invented (and possibly, but not necessarily, patented), society would be better

off in the short run if other firms had instant access to the mousetrap technology and could imitate it and produce it right away. But it is very shortsighted to suppose that the fact that they cannot do so is a barrier to entry or, indeed, is socially detrimental. The profits earned by the mousetrap inventor in the initial period of success represent in part a return on the investment in innovation, an investment which society presumably wishes to encourage. If immediate imitation were to take place, the unnecessary profits on the mousetrap *after the mousetrap has been invented* would indeed be bid away. There would then, however, not be sufficient profits to be earned in the innovation business to induce such innovations in the first place. The right issue is not whether there are barriers to entry into the production of a particular mousetrap, but whether there are barriers to entry into *innovation* in mousetraps. As is always true, the still picture can provide useful publicity but can also give a very misleading idea about the movie.

Conduct and predatory pricing

Thou hast it now: King, Cawdor, Glamis, all,
and I fear,
Thou play'dst most foully for't.

Macbeth, Act III, Scene I

If we think of the conventional structure–conduct–performance paradigm, this discussion has been mostly about structure (an exception being some of the discussion of profits). Unfortunately, it has turned out that many of the structural issues are quite complicated. Is it possible to go about the matter in a different way, by looking at the way an actual monopolist behaves? Are there certain kinds of activity which one can point to as clear evidence of monopolizing behavior?

Put the matter a different way. Aside from the relevance of an examination of conduct to analysis of attempts to monopolize (an offense related to, but different from that of monopolization under the Sherman Act), suppose that we observe a firm with monopoly power; indeed, suppose (what is not the same thing at all, but is a good deal easier to observe) that we observe a firm with a high share of the market. The question naturally arises as to how that share was obtained. The courts have said that it matters whether or not it was obtained by 'superior skill, efficiency and foresight' or 'by conduct honestly industrial, but not economically inevitable'. What kind of conduct ought then to be permitted?

This is a good question for public policy. As I have already indicated, however, I believe that it is awkwardly expressed from the point of view

of economic analysis, even though the issues which must be considered in order to answer it are the same no matter how it is posed. A firm which maintains a large share of the market because of behavior forced on it ('economically inevitable') or solely because of being better ('superior skill, efficiency and foresight') is a firm which does not have monopoly power at all. Monopoly power is the power to maintain a high share and earn supranormal profits *without* being better. A firm which has acquired a large share simply by winning the competitive race through competitive means is not a firm which is a monopoly. The view that it can be a monopoly, but one which is nevertheless to be encouraged stems, I think, from the confusion of high market share with monopoly power which I have already discussed.

Still, whichever way one wants to put it, the question of what kind of conduct is to be encouraged and what prohibited in the course of attaining a large share is an important one. Moreover, the fact that the law can be violated by conduct not in itself illegal but which is 'honestly industrial but not economically inevitable' makes the question of prohibited conduct a difficult one to answer. It also leads to arguments from plaintiffs about all sorts of innocent-appearing and quite possibly actually innocent and competitive conduct.

Is there any rule that one can apply in assessing conduct? I think there are two principles which one can state. The first such principle is that conduct, to be suspect, ought at least to be more restrictive than necessary. The example of the United Shoe Machinery Company which required very long leases and enforced penalty clauses differentially depending on whether or not the customer went to a competitive machine comes immediately to mind. So does the example of Alcoa which bought up power sites far in advance of any use of them (although there the question of distinguishing that behavior from the use of superior foresight at least arises). In both cases, one can say that the conduct involved restrictions to competition which basically had no other purpose. The market could have functioned and the firms been profitable with less restrictive action.

The second principle (and the one the overlooking of which leads to confusion) is that conduct should not be condemned if it is precisely the conduct which competition would lead us to expect. One has to be careful to distinguish between cases in which competition is forcing firms to react and cases in which firms are taking unnecessary action to forestall competition. The competitive model itself points to situations in which firms, faced with competition, will be forced to do certain things or lose business. Firms observed to be doing those things in those situations should not be regarded as monopolizing. They are engaging in conduct which competition makes 'economically inevitable'.

This can be made clearer by considering a leading example, the

analysis of predatory pricing. The notion of a monopolist setting a predatorily low price in order to stifle competitors is a popular one (although less popular among economists than among others). It is certainly fostered by those firms which have to compete against low prices. What can be said on this issue? (Actually quite a lot can be and has been said on this issue and I can do no more than touch on what seem to me to be the really basic points. Recent papers include Areeda and Turner (1975, 1976), Posner (1976), Scherer (1976) and Williamson (1977). The view taken here is closest to that of Areeda and Turner.)

Most economists would at least agree with the following proposition. At any moment in time, there are costs to producing a product which are sunk and costs which can be avoided if the product is not produced. Planning at any moment, a firm should price so that anticipated revenues at least cover the avoidable costs. Deliberate pricing lower than that should be deemed predatory. A firm engaging in that kind of pricing is deliberately earning losses which it could have avoided.

Note, however, that what matters here is the expected revenue and costs as of the time plans are made. In the event it may turn out that demand was considerably weaker than had been expected and products actually make losses. Indeed, after a product is introduced and fails to sell, the best a firm may be able to do is to lower the price of the product to bring in some return on the *then* still avoidable costs. The fact that, viewed after the event, the firm failed to earn a profit on the original investment in developing and producing the product is beside the point. Failures are not necessarily predatory.

Note also, that in assessing what expected revenues and avoidable costs were, one needs to take everything into account. Consider, for example, a firm which bids on a contract to develop a high technology product for the government in the expectation that the experience so gained will be useful in the profitable production of a similar product later on for the commercial market. Such a firm is not necessarily engaging in predatory conduct if it bids for the government contract at a price that does not return its costs in the original development. The revenues from engaging in the development properly include the profits earned later from gaining experience with the technology.

Of course one has to be careful about this sort of thing. A firm which does engage in predatory pricing is presumably doing so because, after competition is forced out, it will earn monopoly profits. Considered as a long-run proposition, the pricing behavior engaged in in the first place was profitable, including in revenues the monopoly profit later earned as a result. Still, it is possible to make sensible judgments about what later effects should and should not properly be included.

The issue over which there is considerable confusion, however, is

that of whether prices which are planned to be remunerative, even considering only direct effects, can also be predatory prices. Imagine a firm charging a relatively high price. Suppose that other firms attempt to enter and produce the same product. Suppose that the first firm, which for some reason is more efficient than the others, drops its price below the costs of the other firms but not below its own costs. Is this 'below-cost pricing' and therefore predatory? Does not it reveal that the initial firm had monopoly power? After all, apparently it had the power 'to set prices and exclude competitors'.

Anybody who is ready to answer 'yes' to those questions should stay after school, because they have missed a good deal of the point of what I have been saying about the analysis of competition and monopoly. Monopoly power is the power to keep prices *high*, earn supranormal profits, and *still* exclude competitors. *Any* firm which is more efficient than its rival always has the power to exclude competitors by setting prices low. That is what competition is supposed to encourage. Further, suppose firms are equally efficient. Then the firm with the largest supply of cash has the power to set prices low and drive out its competitors because it will go out of business last. This is wholly irrelevant. Where competitors are forced out or entry forestalled only by prices being kept low, competition is doing its job. The hallmark of monopoly power is the ability to set *high* prices and earn *high* profits without inducing entry and competitive growth. It is always possible to set *low* prices and earn *low* profits without doing so.

Let me consider two somewhat more specific examples to drive this home. Consider first a firm producing two varieties of the same product. Suppose that it perceives a special demand for one of the varieties and raises the price for it, hoping to earn profits. Suppose it then turns out that others discover a way to make the low-priced variety into the high-priced variety at relatively low cost and that they then begin to purchase the low-priced variety from the original firm, alter it and resell it, undercutting the original firm's price for the high-priced variety. The original firm then finds it cannot maintain the price differential and it readjusts its prices accordingly. The arbitraging firms then find that the profit opportunity disappears and they leave the business. Was the readjustment in prices predatory?

I should hope the answer would clearly be seen to be 'no'. What I have described is exactly what is supposed to happen under competition. A price differential which does not reflect a cost differential is bid away. Had the original firm been able to sustain that differential, there would have been monopoly power. The closing of the differential was in fact 'economically inevitable'.

It is worth commenting further on this example. Note how misleading

it is to look at it only as the arbitraging firms are driven out. By concentrating on the lowering of the price differential, one can pretend one is seeing an attempt to drive out competitors; but that would only make sense if, after they are driven out, it is possible to increase the differential again without bringing them back. That requires some other barrier to competition than merely the lowering of the prices.

Further, competition does not take place in real markets in an impersonal way. The original firm in this example is very likely to understand what is happening and to consider in some detail the costs that the arbitraging firms incur in changing one product into the other. It may even know that it will put them out of the arbitrage business by lowering its price. Even so, it is not committing a predatory act. It is forced, in the example, to lower the price differential or lose business in the high-priced good. What we are observing is competition driving that price differential down to the differential cost of the product, nothing else.

To take a second example, consider the process of innovative competition which I have previously described. In the initial phase, the inventor of the better mousetrap is charging a high price and earning apparently supranormal profits which represent the return on the innovation. Then the imitators come in and bid the price down. *What on earth does bidding the price down mean if it does not mean that the original firm has to lower its price or else lose market share?* Is such a lowering of the price predatory? Surely not. Again, had it been possible to keep the price up without losing share, monopoly power would have been present. Competition will force the innovator to lower the price and, if the innovator's costs are lower than those of the imitators, it will force the price down to the point where the imitators cannot make a profit. Moreover, the innovator may lower its price quite knowingly.

To condemn such conduct as predatory is to condemn exactly the kind of conduct that competition is supposed to foster. If 'economically inevitable' means anything at all, it means this sort of thing. If market share gained by 'superior efficiency' is to mean anything at all, it must mean lower prices by the efficient firm. One can certainly expect cries of outrage and even lawsuits from the others involved, but protection of competition does not mean the protection of individual competitors. Where competitors are being kept out by low prices, competition is doing its job.

That is not to say that low prices cannot form part of an anti-competitive plan. There may be cases in which firms lower prices, thus driving out temporarily less efficient competitors, and in which barriers to entry of some other kind then arise which enable the formerly low-priced firm to become high-priced with impunity. Further, at least in theory, if every attempt to enter is met with aggressive behavior and this

is combined with economies of scale, then some less efficient potential entrants may get the message that there is no room to make profits and the incumbent firm may be able to make profits itself by raising prices; but fully efficient firms (whose entry would clearly benefit society) cannot be deterred in this way unless prices are cut below the incumbent firm's own cost. Moreover, even firms which would become efficient if they stayed and overcame initial scale or other obstacles can only be so deterred if the capital market fails to provide funds in a way commensurate with the true risk involved – the problematic capital barrier to entry already discussed. (Note that *correct* assessment of the high risk involved in overcoming initial inefficiency does not count here even if it makes borrowing difficult.) Even here it seems to me that public policy would be better served by a program of loan subsidies than by trying to discourage firms from choosing the low but remunerative prices that competition is supposed to bring about.

Note further, that where predatory pricing is an issue, a crucial question is whether prices will later be high. Where entrants are merely kept out by low prices and prices cannot be raised without encouraging entry, then competition is working. Competition is supposed to bid prices down. A firm charging high prices, faced with competition, is supposed to lower such prices. Even a firm with 100 per cent of the market, which is only able to maintain that share by so-called 'limit pricing' in which it must keep the prices below the costs of potential entrants, is not by that fact alone engaging in monopoly.

In short, then, while predatory pricing is a possibility, one has to be careful to distinguish it from the behavior competition is supposed to produce. Certainly, in any actual monopoly antitrust case in which most or all of the behavior challenged consists of low rather than high prices, one should be very suspicious.

Conclusion

The message just given relating to predatory pricing is, of course, the message that has run throughout this chapter. In diagnosing monopoly, one has to be careful to distinguish the symptoms from those of competition. Surprisingly, that is not as easy as might be supposed.

I have said relatively little about performance in this regard. By performance, economists generally mean the extent to which a particular market produces the results that one might expect from a competitive market (low profit margin – appropriately measured; efficient means of production employed; and technical progress). It is not always possible to tell when a market's performance is competitive, although it

is usually possible to tell when it is reasonably good. The difficulty with using performance as an index of monopoly power is that, while clearly bad performance would signal the lack of competition, the discretion which monopolists have means that good performance can result even when competition is not structurally present. It is logically true, therefore, that one cannot infer the absence of monopoly power from good performance.

Nevertheless, performance seems to me to be quite a relevant indicator in the following sense. I have tried to show in this chapter how difficult it is to tell monopoly from competition even on grounds of structure and conduct. Doing so is not impossible, but it is easy to become confused and to accept simplistic solutions. Where the performance of a market appears to be good, it seems to me particularly important to be careful about the analysis of structure and conduct. Economists and others ought to approach the public policy problems involved in these areas with a certain humility. Real industries tend to be very complicated. One ought not to tinker with a well-performing industry on the basis of simplistic judgments. The diagnosis of the monopoly disease is sufficiently difficult that one ought not to proceed to surgery without a thorough examination of the patient and a thorough understanding of the medical principles involved.

References

- P. Areeda and D.F. Turner, 'Predatory pricing and related practices under Section 2 of the Sherman Act', *Harvard Law Review*, Vol. 88 (February 1975), pp. 697–733.
- P. Areeda and D.F. Turner, 'Scherer on predatory pricing: a reply', *Harvard Law Review*, Vol. 89 (March 1976), pp. 891–903.
- J.R. Hicks, 'Annual survey of economic theory: the theory of monopoly', *Econometrica*, Vol. 3 (January 1935), pp. 1–20.
- R.A. Posner, *Antitrust Law, an Economic Perspective* (Chicago: University of Chicago Press, 1976), pp. 184–96.
- F.M. Scherer, 'Predatory pricing and the Sherman Act: a comment', *Harvard Law Review*, Vol. 89 (March 1976), pp. 869–90.
- C.C. von Weizsäcker, *Barriers to Entry: A Theoretical Treatment*, (Berlin, Heidelberg, New York: Springer-Verlag, 1980).
- Oliver Williamson, 'Predatory pricing: a strategic and welfare analysis', *Yale Law Journal*, Vol. 87 (December 1977), pp. 284–340.