UNITED ARAB REPUBLIC

THE INSTITUTE OF NATIONAL PLANNING

Memo No 633

COMPETITIVE INDUSTRY

By

Alfred W. Stonier
&
Douglas C. Hague

Reprint from: A Text book of economic theory by A.W.Stonier and D. C. Hague London, Longmans, 1960. Chapter IV

CAIRO
3, MOHAMED MAZHAR - ZAMALEK

Reprint from: A Text book of economic theory , Alfred W. Stonier-Douglas.Hague Ch. v-viii 1960.

(for internal use only)

C H A P T E R VI
COMPETITIVE INDUSTRY

1. DEFINITION OF PURE AND PERFECT COMPETITION

In the previous chapter we discussed the meaning and structure of the various types of curves which we shall need to use in discussing the conditions under which the firm and the industry will be in equilibrium. So far, however, we have merely discussed the various possible shapes of such cost and revenue curves. It is now time to see, using these curves, how the firm reaches equilibrium in various conditions. The simplest case with which to begin is that where there is keen competition. We shall therefore proceed now to analyse conditions for the equilibrium of the firm in a competitive industry. This will not only enable us to see what conditions in such in industry will be like, but will also bring to light in a simple context most of the fundamental features of the equilibrium of the firm. For however competitive or monopolistic the conditions facing a firm happen to be, the main features of the equilibrium of the firm remain essentially the same.

It will be useful, at the outset, to define competitive conditions more carefully than we have done so far. We spoke in Chapter V of pure competition where demand for the product of the individual firm was infinitely elastic, so that the firm could sell all it wished at the existing market price, but was unable to alter the price by its own actions. We saw that in such conditions the average revenue curve of the firm would be a horizontal straight line. There are three fundamental prerequisites for the existence of pure competition between producers. These are as follows.

(a) Large Numbers

The first condition for pure competition in an industry is that there must be a large number of firms in the industry. This is essential, because only when there are many firms is an industry can each firm be sure that any action on its own part will have no noticeable effects on the price and output of the whole industry. If there are many firms in an industry, any one of them can increase or decrease its own output quite substantially without needing to fear that there may be a violent change in the price of the product of the industry as a result. The firm will be producing such a small proportion of the total output of the commodity that even a large change in its own output will have a negligible effect on the output and the price of the industry as a whole. The individual firm has to take the price of the product as given and unalterable.

(b) Homogeneous Products

Second, each of the firms in a 'purely' competitive industry must be making a product which is accepted by customers as being identical, or homogeneous, with that made by all the other producers in the industry. This ensures that no entrepreneur can put his price up above the general level. If he did do so, consumers could buy the same good from other producers at a lower price. Best-grade cotton is best-grade cotton whichever planter grows it, and one does not find infividual cotton planters advertising their product as better than that grown by their next-door neighbour if both are of equal quality. When the goods are homogeneous in this way a uniform price must rule throughout the market.

It is, however, important to realise that it is the consumer who decides whether or not two products are homogeneous.

It is only if buyers agree that two producers are making exactly the same good that their prices will be the same. If buyers find real or imagined differences between the two products, their prices are bound to differ, however alike the two articles really are. For example, if a customer wrongly believes that the 'Co-op' sells goods which are inferior to those supplied by 'Quality Groceries' he may pay the latter 3s. 6d. for a good which he could have got from the former for 2s., even though the goods are identical. Again, Mr. Smith may prefer to buy his gardening implements from Mr. Jones, because they both belong to the same club or go to the same church, even though Mr. Jones changes higher prices for the same goods than Mr. Williams farther down the street. If buyers behave in this way, competition cannot be pure, for the products are not homogeneous in the eyes of the consumers.

These two assumptions between them ensure that the firm's average revenue curve shall be horizontal. The fact that there are many producers prevents the individual firm from exerting any influence on price. The fact that products are homogeneous means that buyers do not regard one producer's product as 'better' than another's and prevents price differences from emerging.

It is probably worth pointing out here that we have so far assumed, and shall continue to assume until further notice, that there is always pure competition between buyers. We have taken it for granted that the total number of buyers is very large, and that each one takes so small a proportion of the total sales of any good by his own actions. Buyers must therefore take prices as given, even where a monopolistic seller can alter prices to suit himself. The fact that in this chapter we are assuming perfect competition between both buyers and sellers is important and should be borne in mind.

(c) Free Entry

The third fundamental condition which must be fulfilled if there is to be pure competition in an industry is that anyone who wishes to enter the industry must be allowed to do so. It is difficult, however, to be quite sure what is and what is not a restriction on entry into an industry. For example, in the road passenger transport industry it is desirable that every omnibus or coach should measure up to certain required standards of safety. A refusal to allow unsafe buses into the industry would not usually be regarded as a restriction on entry. Rather would it be accepted as an essential condition for the well-being of passengers, transport workers and other road users.

Nevertheless, whilst border-line cases of this kind may well be important in some circumstances, it will, as a rule, be possible to decide quite quickly whether or not there is restriction on entry into any industry. A monopolistic association, for example, will refuse entry into an industry to anyone whom it does not wish to allow into the industry. If it is able to enforce its decisions, this will mean that the number of firms in an industry can be kept at the existing number, or even reduced if some firms leave the industry, while prices are maintained at a high level. Our third prerequisite for pure competition thus ensures that the number of firms in the competitive industry can always remain large because new blood is always allowed to enter.

These three conditions, large numbers of firms, homogeneous products and free entry, between them ensure that there is pure competition in the sense that there is competition which is completely free from any monopoly elements. Between them they ensure that the average revenue curve of each of the

individual firms in the competitive industry is a horizontal straight line. One can,however, distinguish also between pure competition, which we have just defined, and 'perfect' competition - a concept frequently used by economists. For there to be 'perfect' competition, it is necessary to make some additional assumptions. In particular, it is necessary to assume that there is perfect knowledge on the part of all buyers and of all sellers about conditions in the market. In addition it is usual to assume complete mobility of factors of production between industries. It is also convenient when discussing perfect competition to make the assumption that all producers work sufficiently close to each other for there to be no transport costs. Strictly speaking, tow identical goods are not homogeneous in the eccnomic sense if they are not in the same place. For example, I may have the choice between buying identical kettles from Smith in my village or from Brown in the next village. But if the bus fare is 6d. each way, Smith may charge me 1s. extra because he knows it would cost me 1s. to fetch the same sort of kettle from the next village. We shall assume that differences caused by transport costs do not exist, otherwise prices for the same physical good would differ. These additional assumptions are not essential if one merely desires to show what conditions must be fulfilled if monopoly on the selling side is to be absent. But they are useful in building up a phypothetical model of a competitive industry. We shall therefore discuss the conditions in an industry where there is 'perfect' rather than 'pure' competition. This is our next task.

2. NORMAL PROFIT

Two essential conditions must be fulfilled if there is to be equilibrium in a perfectly competitive indusyry. First, each and every individual firm must be in equilibrium. This will happen when each firm in the industry is earning maximum profits, by equating marginal revenue with marginal cost. Second, the industry as a whole must be in equilibrium. This will occur when there is no tendency for firms either to enter or to leave the industry, which will only happen when all the entrepreneurs in the industry are earning enough money to induce them to stay in the industry, and when no enterpreneur outside the industry thinks that the he could earn enough money, were he to enter it, to make the move worth while.

If these conditions are fulfilled, and there is no movement of firms either into or out of the industry, it is clear what the entrepreneurs of all the firm which are in the industry must be earning at least enough money to induce them to remain in it. In technical economic language, we describe this situation by saying that every entrepreneur in the industry is at least earning normal profits. 'Normal' profits, for an entrepreneur in any industry, are those profits which are just sufficient to induce him to stay in the industry. It follows, on the other hand, that if an industry is in equilibrium, with no movement in or out, no-one outside the industry foresees the possibility of being able to earn 'normal' profits if he were to enter the industry.

We shall need to make considerable use of the idea of 'normal' profits, and it will be useful, in this chapter, to simplify the problem by making two not entirely realistic assumptions about it. Let us assume that all the entrepreneurs

in the industry we are studying would all prove to be of identical efficiency if they left it and went to another industry. On this assumption, 'normal' profits will be identical for every enterpreneur in the industry we are studying. But let us also assume (unless an explicit statement to the contrary is made) that the entrepreneurs are not all of equal efficiency within the industry and that some can therefore earn more money than others. Some may earn 'supernormal' profits. These assumptions imply that there is a general rate of 'normal' profits for the whole industry-that all entrepreneurs' earnings must fall to exactly the same level before they will leave the industry-but that within the industry some entrepreneurs earn more than others. These two assumptions are not entirely consistent but they will enable us to simplify the analysis in this chapter. They will be abandoned in a later chapter. 1

It follows from the foregoing analysis that if, for any reason, profits of all entrepreneurs in any industry rise above 'normal' (if they are 'supernormal'), there will be a tendency for the number of firms in that industry to increase. For we may assume that entrepreneurs outside the industry will expect to be able to earn at least 'normal' profits if they enter. If, on the other hand, profits for everyone fall below 'normal' (if they are 'subnormal'), there will be a fall in the number of firms in the industry. Some firms will be forced out by bankruptcy and their entrepreneurs will go in search of 'normal' profits elsewhere. We can say, therefore, that the industry as a whole and all the individual firms in it will be in equilibrium- in 'full equilibrium'as it is often called - when all firms are maximising profits and when there is no tendency for firms to enter or leave the industry. In such conditions all the firms in the industry will be earning at least normal profits.

The fact that we have introduced normal profits into our analysis means that we must make a slight alteration in the content of our average cost curve. If a firm is to remain in the industry, its decision to do so will depend, not on whether it is covering average total cost as we have hitherto defined it, but on whether it is also earning at least normal

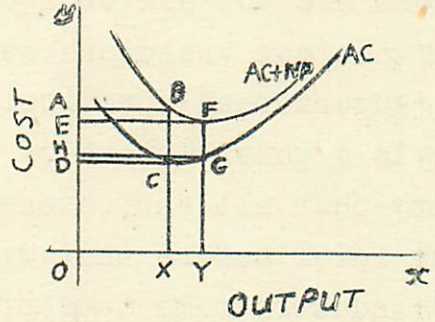


FIGURE 51

profits. So, in order to be able to construct an analysis which enables us to decide easily whether or not firms will be willing to stay in an industry at their current level of earnings, it will be useful to include 'normal' profits in average costs. In future, the average cost curves which we shall draw will include such 'normal' profits. This should be remembered, because it will not be convenient to point it out specifically on every occasion when an average cost curve is drawn. The introduction of normal profits gives an additional reason for us to expect average cost curves to slope downwards over low ranges of output. We shall assume that each entrepreneur must earn a certain amount of money if he is to stay in the industry, and it is usual to assume that this sum of money representing normal profits if independent of output. We shall assume that an entrepreneur must earn fixed sum of money if he is to stay in the industry, and that he wants that fixed amount of money whatever output he is to produce. Because normal profits represent a fixed amount of money, it follows that, as output rises, normal profits calculated per unit of output fall. For the fixed sum of money representing normal profits

is spread over a progressively larger number of unit of output as production rises. Here is our additional reason for expecting average cost curves to fall over low ranges of output.

In Figure 51 we have drawn an average cost curve excluding normal profits (AC) and another including them (AC + NP). It will be seen that as output rises, the vertical distance between the two curves steadily falls. Normal profits per unit of output decline progressively. For example, when output is OX, normal profit per nuit of output is BC, when output is OY it is FG. It should be noted that the areas of any rectangles showing normal profits such as the rectangles ABCD and EFGH are the same. This is inevitable if, as we are assuming, normal profits represent a given sum of money.

If average cost includes normal profits in this way, it is clear that a firm will be earning normal profits when average cost (including normal profits) is equal to price. Now, we have

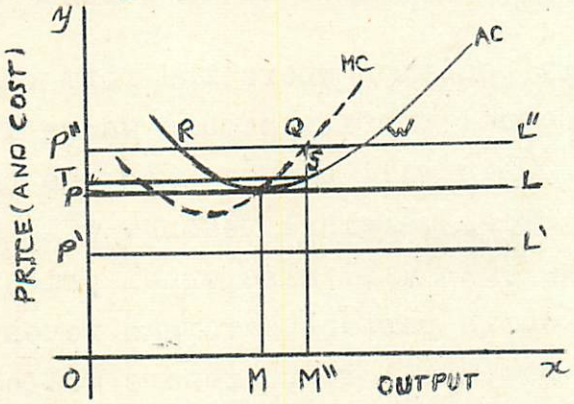


FIGURE 51

seen in Chapter V that it is reasonable to assume that all average cost curves are roughly U-shaped. We alsoknow that in perfect competition each firm's average revenue curve is a hori-zontal straight line. Thus, in perfect competition the only situation in which the firm can be in equilibrium and earning normal profits is when the average cost curve is tangential to the average revenue curve. Only then can average

revenue equal price, and the firm cover all its costs and just earn normal profits. This shown in Figure 25.

In Figure 25 it is clear that the firm cannot both be in equilibrium and be earning normal profits for any position of the average revenue curve below PL. For example, with the average revenue curve P'L' the firm must earn less than normal profits whatever its output, for at no output is average cost (including normal profits) equal to price. On the other hand, if the average revenue curve is above PL, as with P"L", it is possible for average cost to equal price (in this case at R and W). But at neither R nor W would the firm be in equilibrium. For the firm will only be in equilibrium when it is equating marginal revenue and marginal cost. With the average revenue curve P"L", this will happen when the firm produces the output OM and sells it at the price OP". At this level of output it will be in equilibrium but will be earning 'supernormal' profits, equal to the area P"OST.

The only position where the firm can be in equilibrium and only earn normal profits occurs where it is producing the output OM. This will happen when the price of the industry's product is OP. Then marginal revenue will be equal to marginal cost but average cost will also equal price (average revenue). Since competition is perfect, average revenue will, of course, equal marginal revenue. It therefore follows that if a perfectly competitive industry is to be in full equilibrium, with all firms not only earning maximum profits (equating marginal revenue with marginal cost) but in fact earning normal profits with average cost (including normal profit) equal to price, the average cost curve of each firm must be tangent to its average revenue curve.

firms, and since competition is perfect it is also the marginal revenue curve.

All firms have identical costs, so each will maximise profits if it produces the output OM and sells it at the perfectly competitive price of £ OP. Since average cost is only £ MR at the output OM, and price is £ OP, each firm earns £PLRP of 'supernormal' profits. Both firms in Figure 53 are in equilibrium, since both are earning maximum profits. But the industry is not in equilibrium, for all firms have identical cost curves. They will thus all be producing the same output and earning the same 'supernormal' profits-£ PLRP'. Since no firm can creat new fixed equipment in the short run, no new firms can enter the industry, except in the long run. In the long run new firms will enter the industry and, having identical costs with the existing firms, will compete away the supernormal profits. In the short run, however, the number of firms in the industry is too small for there to be 'full equilibrium', and all firms earn supernormal profits.

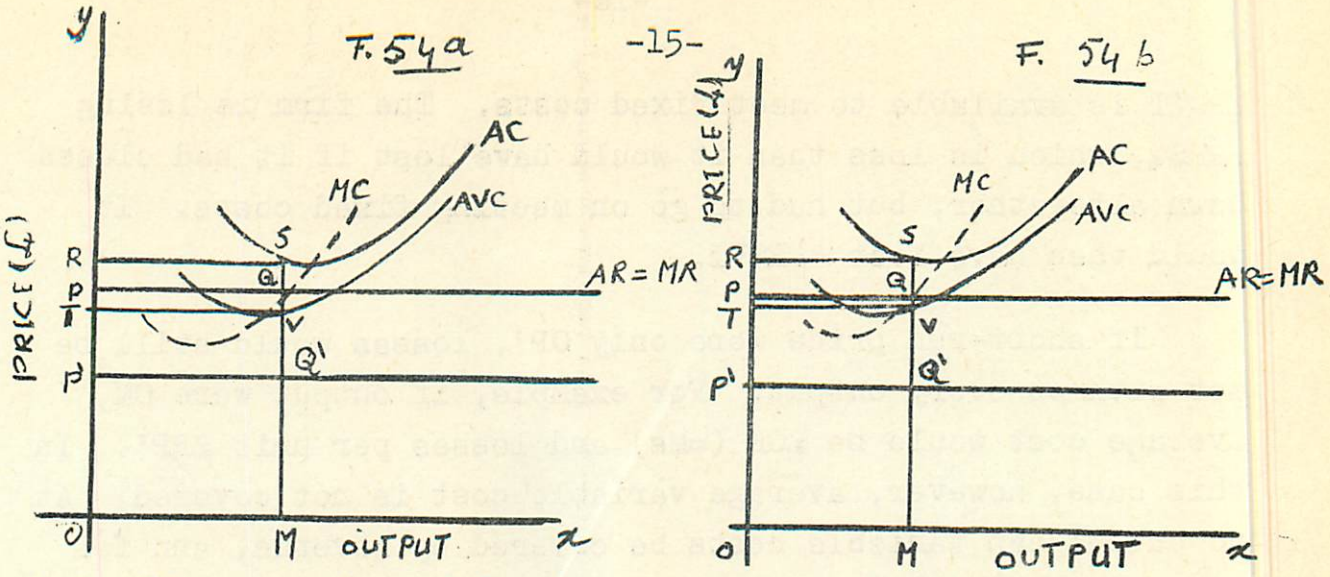
It will be noted that in both firms in Figure 53 marginal revenue and marginal cost will be equal at L, and that it is by reference to this point that the profit-maximising output will be determined. But marginal revenue and marginal cost will also be equal at point T. Why is it that the firms are not in equilibrium at this level? This becomes clear on examination of Figures 53a and 53b. At T marginal cost has only just become equal to marginal revenue and has previously been greater. If output were fixed at this point, the firm would be earning minimum and not maximum profits. We may therefore state the conditions necessary for equilibrium of the firm in perfect competition more explicitly. The marginal cost curve must cut the marginal revenue curve from below at the point of equilibrium. A firm can never be earning maximum profits unless this happens.

If price were £OP"instead of £OP, the firms would be in equilibrium when producing the output OM'. They would just be covering their costs and would be earning normal profits. This condition can be seen visually from the fact that the average cost curve is tangent to the average revenue curve. The situation thus represents full equilibrium in the industry, even in the short run. The number of firms is just large enough to ensure that no-one earns supernormal profits.

If, on the other hand, price happened to be OP"', the firms in the industry would be in equilibrium producing the output OM". All would be losing money, losses being equal to P'R'R"P"'. Thus the 'maximum' profit a firm can earn is a loss of £ P'R'R"P"' . This is the smallest loss the firms can make if they are to produce at all. It follows that whilst the firms can be in short-run equilibrium, they will not be in long-run equilibrium. In the long run firms will leave the industry until those remaining just earn normal profits.

The obvious question to ask therefore is: 'Does it pay a firm which is losing money to stay in the industry?' The answer is that it all depends on the length of time which is being considered. In the long ruun, if the efficiency of the firm remains unaltered, it should leave the industry altogether. By definition, if the entrepreneur is not earning normal profits he will prefer to leave the industry in the long run.

In the short run he may prefer to stay in business. For by our definition of the short period he cannot, within such a period, alter the fixed capital of his firm, and this fact will influence his decision. Since the 'fixed' costs of



of the firm must be met in the short run, even if the firm colses (unless it is to go bankrupt altogether), it will pay the entrepreneur to remain in production if by doing so he earns anything which will help him to cover his 'fixed' costs. This can be seen from the firms in Figure 54. We still assume identical costs. The firms will remain in business solong as they can cover their variable costs. Since variable costs can be avoided, even in the short run, the firms can refuse to employ any variable factors if they do not wish to do so. It will therefore pay entrepreneurs to take on variable factors only if they can pay for them and still have something left over to help to meet fixed costs. This will, however, be no more than making the best of a bad job in the short run, and will not provide a long-run solution to their troubles. That can only come from a reduction in costs or a rise in price.

Looking at the short-run problem in greater detail, we can see that in each firm in Figure 54 the 'loss minimising' output is OM, for price is £OP per unit of output. Price (average revenue) therefore exceeds average variable cost MV (as shown on the average variable cost curve AVC) by ₠v. This means that it will be worth while for the firm to continue running in the short run. For some fixed costs can be covered after variable costs have been met. In this case, for example,

£PQVT is available to meet fixed costs. The firm is losing £PRSQ, which is less than it would have lost if it had closed down altogether, but had to go on meeting fixed costs. It would then have lost £RSVT.

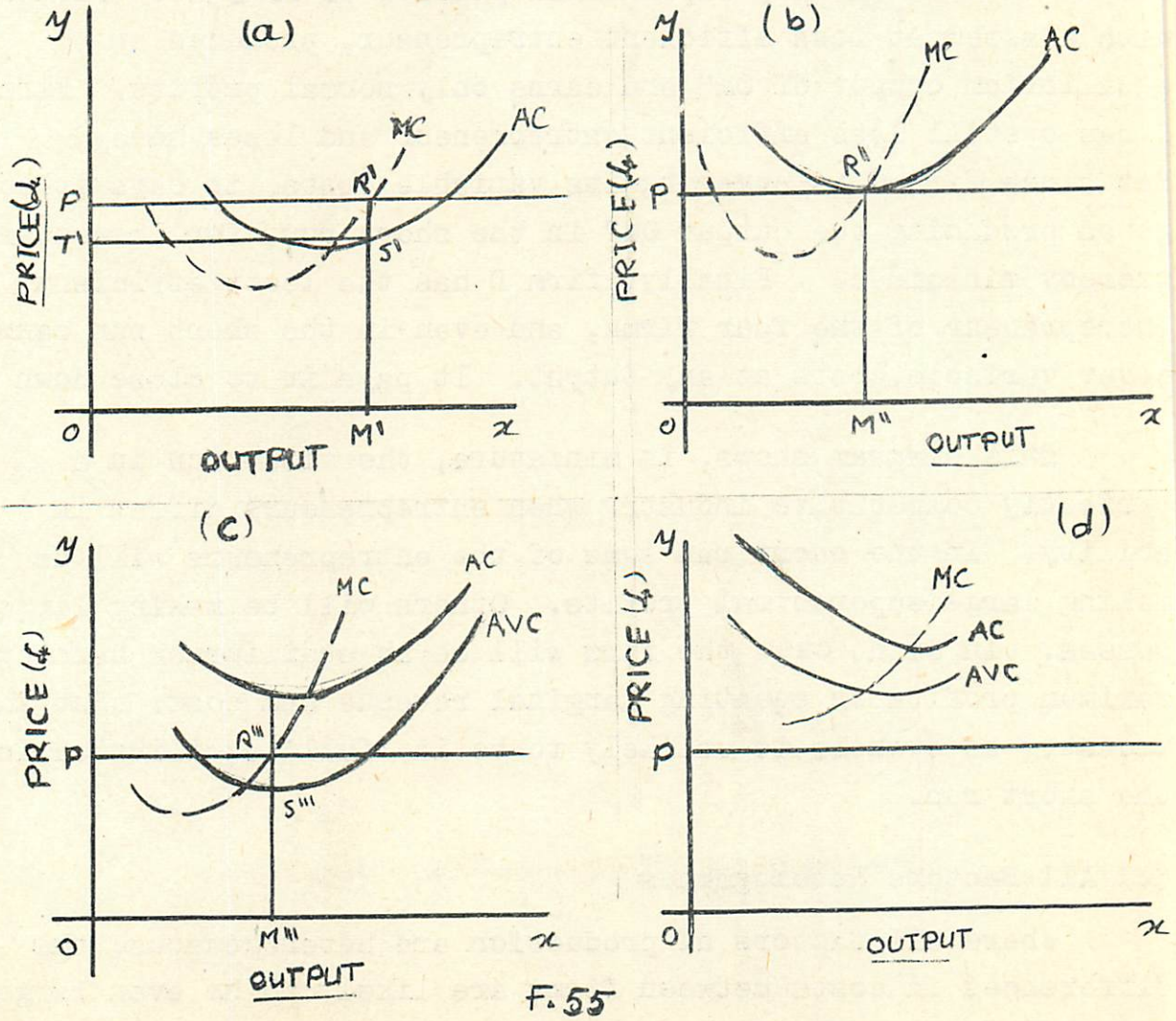
If short-run price were only OP', losses would still be met with at every output. For example, if output were OM, average cost would be £OR (=Ms) and losses per unit £RP'. In this case, however, average variable cost is not covered. At no output can variable costs be covered by revenue, and far from helping to meet fixed costs, the fact that the firm remains in production is merely making things worse. For example, at output OM the firm is losing an additional £P'O'VT by refusing to close down altogether. It will therefore pay the firm to stop producing anything at all, even in the short run, though it may be possible to reopen later.

What can we conclude from this analysis? We can see that, in the short run, it is possible for firms to bring themselves into a position of equilibrium where they are producing that output which maximises profits or minimises losses, even though in some cases this may mean that production stops altogther. One is entitled to assume that it will not be difficult for the firm to do this even in a short period of time. But it will be very much less likely that the industry will be able to bring itself into a 'full' equilibrium in the short run. For such a 'full' equilibrium can only occur when all firms are earning normal profits. A little reflection will be sufficient to convince one that it is only by accident-only if there happens to be just the right number of firms-that the whole industry will be in equilibrium in the short run. It is much more proable that there will have to be a long-run adjustment before the number of firms in the industry can change and full

equilibrium be brought about. It is important to realise that we have shown two firms only in our diagrams, simply for the sake of convenience For, since all factors are homogeneous, the situation will be the same in every firm in the industry. But it should not be forgotten that the analysis only applies to an industry where there is a large number of firms producing a homogenous product.

(b) Entrepreneurs Heterogeneous, other Factors Homogeneous

If entrepreneurs are not all identical in efficiency, then costs can differ between firms even though all factor



F.55

prices are the same to all firms and all factors except entrepreneurs are homogeneous. The more efficient entrepreneurs will be able to produce more efficiently than the others, and their firms' costs will therefore be lower. Different firms will be producing different outputs at different cost, even though all firms make the same product and sell it at the same price. The kind of situation which is likely to exist in the short run is shown in Figure 55.

Firm A has the most efficient entrepreneur of the four firms shown, and it is in short-run equilibrium producing the output OM' and earning supernormal profits of PR'S'T'. Firm B, with a somewhat less efficient entrepreneur, produces an equilibrium output of OM" and earns only normal profits. Firm C has a still less efficient entrepreneur and loses money. But since firm C is covering its variable costs, it pays it to go on producing the output OM" in the short run, for losses are thereby minimised. Finally, firm D has the least efficient entrepreneur of the four firms, and even in the short run cannot cover variable costs at any output. It pays it to close down

This diagram shows, in miniature, the situation in a perfectly competitive industry when entrepreneurs differ in ability. In the short run some of the entrepreneurs will be making large supernormal profits. Others will be making large losses. In each case the firm will be in equilibrium earning maximum profits by equating marginal revenue and cost. But the industry as a whole is unlikely to be in 'full' equilibrium in the short run.

(c) All Factors Heterogenous

Where all factors of production are heterogeneous, the differences in costs between firms are likely to be even larger than where only entrepreneurs are of differing efficiency.

The same kind of situation as shown in Figure 55 will occur, but the differences between the costs of the various firms are likely to be greater. For the firms which can obtain labour which is more efficient relatively to its wage will have a cost advantage. The more efficient are the factors a firm is using, the greater is profit will be compared with those earned by other firms.

The general conclusion, then, still holds. Each individual firm will be able to reach an equilibrium profit-maximising position in the short run, though it may (in very bad conditions) have to close down entirely. But there is no reason why the industry as a whole should be in full equilibrium in the short run. As we have seen, most firms will usually be earning either supernormal or subnormal profits in the short run. Only in the long run is equilibrium in the industry as a whole likely to come about-by an increase or a decrease in the number of firms in the industry.

4. LONG-RUN EQUILIBRIUM OF THE FIRM AND INDUSTRY

What, then, is likely to happen in a perfectly comptitive industry in the long run? Because we are assuming that there is perfect competition, there will be no restrictions on entry into the industry and all factors of production will be perfectly mobile. It is therefore reasonable to argue that in the long run firms will enter or leave the industry until no firm outside the industry thinks it could earn normal profits if it were to enter the industry. Let us again consider what will happen in our three sets of conditions. We shall now need to use long-run cost curves instead of short-run ones, but this is the only difference in the analysis.

(a) All Factors of Production Homogeneous

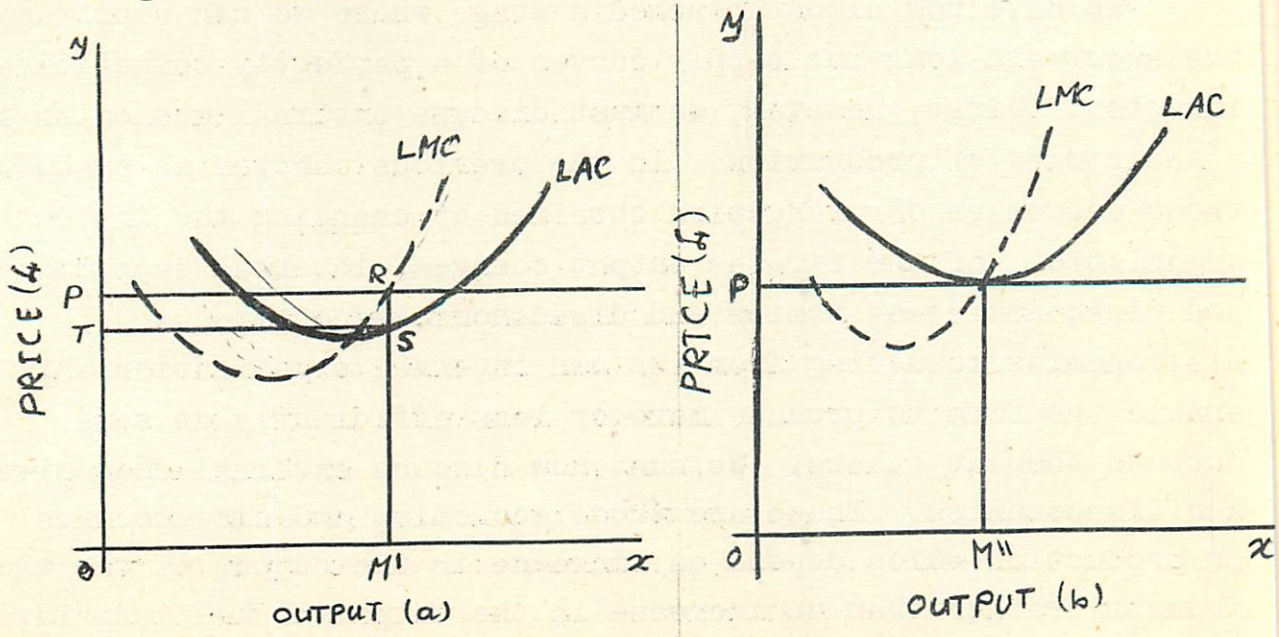
Where all factors of production are homogeneous and each factor has a given and uniform price, it is not difficult to see that the industry will be in 'full' equilibrium in the long run. Each firm will adjust its output so as to equate marginal revenue with marginal cost. But in the long run each firm will also be in the position shown in Figure 56b, where average cost, average revenue, marginal cost and marginal revenue are all equal to each other and to price. This is the 'full equilibrium' position. It will be reached because, since all firms have identical cost curves and since there is free entry and perfect mobility, firms will enter or leave the industry until all are earning normal profits. And because costs are identical, if any one firm is earning normal profits all firms will be. In perfect competition, with all factors homogeneous, each firm and the industry as a whole will be in full equilibrium where marginal revenue = marginal cost =average cost = average revenue (price).

It is interesting to note that in full equilibrium each firm will be producing the 'optimum' output where average cost is at a minimum. This is advantageous to consumers since the product in question is being produced in the cheapest possible way without any firm making a loss. But all firms can only be producing the optimum output if all have the same minimum point to their cost curves, and this is unlikely to happen if all factors are not homogeneous. So we must now turn to the second situation where not all factors are homogeneous.

(b) Entrepreneurs Heterogeneous, all other Factors Homogeneous

As we have seen where entrepreneurs are heterogeneous, some firms will be albe to produce the same output at lawer cost than others. This means that even in the long run it is conceivable that some firms will be able to earn 'supernormal' profits. Let us consider the two firms shown in Figure 56, firm A in Figure 56a and firm B in 56b.

Firm A has a more efficient entrepreneur, and therefore has lower costs than firm B. Let us assume that in the long run firm B is just efficient enough to remain in the industry and to earn normal orofits. We could describe it as a 'marginal' firm. It is 'on the margin of profitability' because any fall in price would, in the long run, send it out of the industry. Since firm A has a more efficient entrepreneur, it is able to earn supernormal profits of £PRST even in the long run. It can therefore be described as an 'intra-marginal' firm. The only way in which firm A could be forced to earn normal profits in the long run would be for there to be an influx of very efficient



F.56

producers able to compete more effectively with firm A than firms like B. This eventuality is, of course, a possible one, but it is not necessarily a likely one, and it would, of course, drive B out of the industry. It therefore follows that even in the long run some firms will be able to earn supernormal profits if they have more efficient entrepreneurs than other firms.

(c) All Factors Heterogeneous

The same result will occur where all factors are heterogeneous. The firms with the most efficient factors will be able to earn supernormal profits even when 'marginal' firms are only just earning normal profits. It is important to bear in mind that this whole analysis is based on the assumption that normal profits are the same for all entrepreneurs-and this need not be the case.

5. EXTERNAL ECONOMIES AND DISECONOMIES

We have now almost reached a stage where we can construct the short-and long-run supply curves of a perfectly competitive industry. First, however, we must discuss external economies and diseconomies of production. In the previous chapter we discussed those economies of production obtained by changing the internal organiztion of the firm as output changes, internal economies and diseconomies-economies and diseconomies-economies and diseconomies resulting from its own internal organisation which enable the firm to produce more or less efficiently at some outputs than at others. We must now discuss external economies and diseconomies. These are those economies and diseconomies in production which depend on increase in the output of the whole industry rather than on increase in the output of the individual firm.

External economies occur where an increase in the size of an industry leads to lower costs for the individual firms composing the industry. For example, in the coal-mining industry, the fact that, in pumping water from its own workings, a mine also pumps water from the workings of other mines, means that the more pits there are in the area, the drier each pit will be. Another very important type of external economy is to be found where the efficient development of the industry depends greatly on the interchange of technical information between firms. In such a case it is obvious that the larger the industry is, the more feasible and worthwhile it will be set up large-scale information services and to publish such things as trade newspapers and magazines.

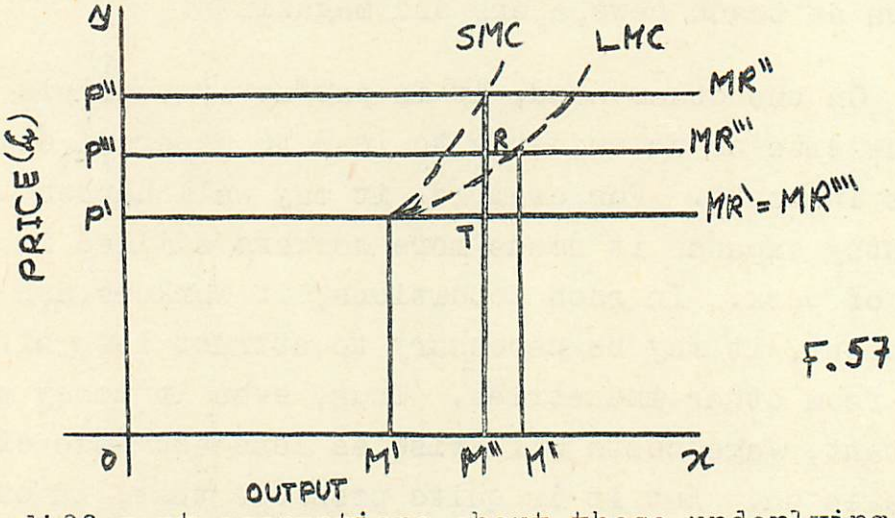
On the other hand, it is perfectly possible for a growth in the size of an industry to lead to external diseconomies-to rises in costs. For example, it may well happen that if an industry expands it needs more workers skilled in a particular kind of work. In such conditions, if workers are not all equally efficient, it may be necessary to attract less efficient workers away from other industries. Thus, even if money wages remain constant, wage costs will rise as less and less efficient labour is taken on. But it is quite possible that, in order to attract labour away from other industries, wages, may have to be raised for the particular type of labour needed. This will be a diseconomy of increased production which is external to the individual firm--the increased size of the industry as a whole has raised the costs of the individual firm.

External economic and diseconomies need to be brought into our analysis because we are about to consider the nature of the supply curve of the perfectly competitive industry. The shape of such a supply curve will depend, in part at least,

whether external economies or diseconomies arise as an industry increases or decreases the scale of its operation.

VI THE SUPPLY CURVE OF THE PERFECTLY COMPETITIVE INDUSTRY

Our main aim in this chapter and the preceding one has been to bring us to a position where we can build up a supply curve for a perfectly competitive industry. This is an essential step before we return to a discussion of the relationships between demand and qupply, outlined in Chapter I. The shape of a competitive supply curve will not, however, always be the same. It will depend on the production conditions underlying it. Let us therefore consider what the shape of the supply curve will be,



with various different assumptions about these underlying conditions. The simplest assumption we can make is that all factors of production are homogeneous and are in perfectly elastic supply to the industry. We shall continue to assume throughout that there is perfect competition in the factor market so that the prices of all units of each factor are the same.

We can consider the supply curve of the industry in three different situations. First, we may see what it will be in

the short run when only is the number of firms given, but also the size of the individual firm cannot be altered. Second, we shall consider a long-run situation where we shall assume that the individual firm is able to alter the scale of it operations, but where, for the moment, we shall assume that the number of firms remains constant. Third, we shall consider a long-run situation where we shall assume that both the number of firms and the scale of operations in the individual firm are able to change.

Let us now assume that the initial situation is one where both the industry and firm are in full equilibrium with each firm producing at the minimum point of both is short-and long-run cost curve. This situation can be seen in Figure 57.

All firms in the industry will have identical cost curves because factors are homogeneous, and will therefore originally be in long-run equilibrium, where they are producing an output of OM' and selling it at a price of £OP'. Marginal revenue, shown by the curve P' - MR', is equal to both short-and long-run marginal cost for all firms in the industry, and all these firms are earning normal profits. Since competition is perfect, marginal revenue will esual average revenue (price).

In order to discover what the shape of the industry's supply curve is, let us assume that the price of the product rises to OP'' and the marginal revenue curve rises to P'' - MR''. In the short run it is apparent that the supply curve of the individual firm is, in perfect competition, its short-run marginal cost curve SMC. The only way in which any firm can produce more in the short run is by expanding output along its short-run marginal cost curve. In fact, if price rises, as in Figure 57, to OP'',

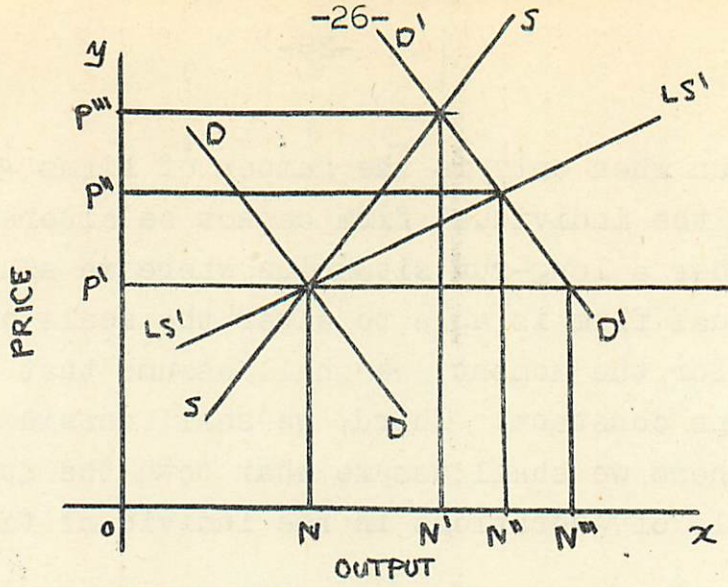


FIGURE 58

marginal revenue will equal marginal cost at the output OM". Since all firms are identical (on our present assumption), the total increase in the output of the industry as price rises to OP" will be M'M" multiplied by the number of firms in the industry. Or, to put it another way, the supply curve of the industry will be a lateral summation of the supply curves of the individual firms. In figure 58 it has been assumed that the industry consists of ten identical firms with the cost curves shown in Figure 57. The short-run supply curve of the industry is the curve SS. It shows what when price rises from OP' to OP", because the demand curve shifts from DD to D'D', output rises from ON to ON' - by NN' - which is ten times M'M" in Figrue 57. (The distance M'M" in figure 57 is not one-tenth of the distance NN' in Figure 58, because the scale on the x-axis is greater in Figure 57 than in Figure 58.)

It is clear that the short-run supply curve of a competitive industry will always slope upwards, since the shrot-run marginal cost curves of the individual firms will always slope upwards too. The steepness of the industry's short-run supply curve will depend on the slope of the cost curves of the individual firms. Where the cost structures of all the individual firms in the industry are identical, the elasticity of the industy's

supply curve will be equal to the elasticity of the separate firms' marginal cost curves. The only general rule one can lay down is that the industry's short-run supply curve must slope upwards to the right.

We must now consider what will be the industry's long-run supply curve, assuming first of all that the number of firms remains constant. If this happens, the long-run equilibrium situation in Figure 57 will be where output is OM"' and price Op"'. The long-run marginal cost curve of the individual firm (L M C in Figure 57) will be less steep than the short-run one, and output will be able to expand more than in the short run (by M'M"' instead of by M'M"). Because output is able to expand, price falls-to OP"'. The long-period supply curve of the industry (with the number of firms fixed) will be the curve L S' in Figure 58. The increase in the output of the industry is now NN" compared with the original output of ON. NN" is ten times M'M"' in Figure 57. This particular type of long-run supply curve thus slopes upwards to the right-though less steeply than the short-run supply curve. It slopes upwards because the long-run marginal cost curves of the individual firms on which it depends, will also slope upwards. It may be noted that in this new long-run situation, with price OP"', all the firms in the industry will be earning abnormal profits, as shown in Figure 57. This can happen, even in the long run, because the number of firms in the industry is assumed to be constant.

Finally, we must consider the more usual long-run situation where not only can the size of all firms in the industry alter, but it is also assumed that the number of firms can change. We assume, in fact, that there is free entry into the industry, and that new firms continue to enter until no abnormal profits

are being earned. In the short run a firm in this condition will produce, in Figure 57, the output OM" at the price OP". But in the long run the individual firms, shown in Figure 57, will be in exactly the same position as in the original equilibrium. Output will be OM' and price OP'. Marginal revenue will be shown by the curve P' - MR"" which is identical with the curve P' - MR'. New firms will have entered the industry and all abnormal profits will have been competed away. The output of each firm is once again OM', but, because there are now more firms, the output of the industry has increased. This can be seen from figure 58. The long-run supply curve, where the number of firms can alter, is the curve P'- LS", and it will be seen that, in equilibrium, the output of the industry is now ON"'. This long-run supply curve is horizontal because new firms have entered the industry-firms which are identical with those already there-and marginal cost in each firm has returned to the original level. Provided that all factors are homogeneous, the long-run supply curve of the industry (with free entry) will be a horizontal straight line and supply price will be the same whatever the output. This is the simplest possible kind of 'free entry' long-run supply curve, and reprexsents a first approximation to reality.

It is unlikely that in practice all factors will be homogeneous. It is much more likely that entrepreneurs, at least, will be heterogeneous. If factors other than entrepreneurs are heterogeneous, it is reasonable to look upon this as an external diseconomy of the industry. All firms may have to hire some less efficient factors if the size of the industry will raise the costs of all firms. For even if all units of factors continue to earn the same money reward, the less efficient will produce a smaller output than the more efficient, and the cost

of such output will rise. If entrepreneurs are heterogeneous, however, there will be no effect on the costs of firms previously in the industry if an expansion brings in new entrepreneurs. It will only be the new firms which have higher costs. This situation is therefore not the same as where heterogeneity of other factors causes external diseconomies. Let us now consider what will be the shape of the supply curve of the industry where entreprenurs are heterogeneous, but all other factors are homogeneous.

So far as the short-run supply curve is concerned, the fact that entrepreneurs are heterogeneous will make little difference. It will still represent a lateral summation of the short-run marginal cost curves of the individual firms. These marginal cost curves will, of course, now be different for each firms, so that the process of summing them will be more complicated than where factors are homogeneous. But, since the number of firms is fixed in the short run, and since the short-run , marginal cost curve of the firm always rise too. Whether or not factors of production are homogeneous, the short-run supply curve is bound to slope upwards to the right, but heterogeneous factors it will slope upwards rather more steeply.

In the long run the number of firms will be able to increase, assuming that there is free entry. But since the new firms which enter the industry in the long run, in order to raise output, will be run by less efficient entrepreneurs, their costs will be higher at each level of output than in existing firms. For long-run equilibrium to occur at a greater output, the price of the industry's product must have risen. Only in that way can less efficient firms be attracted into the industry (and be maintained there).. The rise in price needed to produce a given increase in output will be amaller than in the short run. This is partly

because in the long run firms will be aboe to produce an increased output more efficiently than in the short run. Partly, however, it is because in the long run, as the number of firms increase, existing individual firms will be able to reduce their output again and so reduce their marginal costs. The supply curve of the industry, with heterogeneous entrepreneurs and all other factors homogeneous, will therefore slope upwards, whilst with all factors homogeneous it will be horizontal. But it is likely to slope upwards rather less steeply in the long run than it does in the short run.

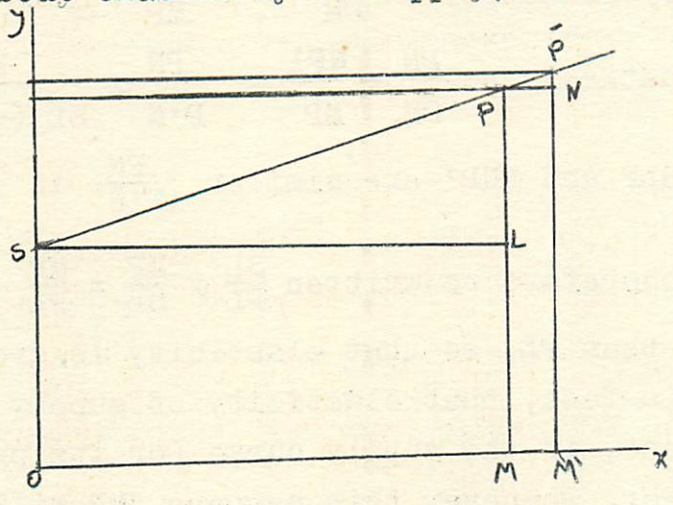
Finally, we must consider what will be the shape of the industry's supply curve where there are external economies and diseconomies. We may here confine out attention to the long-run supply curve. External economies and diseconomies can only be important in the long run, for only in the long run can the size of the industry alter and any repercussions of this on costs of production occur.

We have seen that if all factors of production are homogeneous and in perfectly elastic supply, the long-run supply curve of the industry will be horizontal. If in such conditions there are external economies, the long-run supply curve will slope down-words. If, on the other hand, there are external diseconomies, the supply curve will slope upwards. We have also seen that where entrepreneurship is not homogeneous in a perfectly competitive industry, but all other factors are homogeneous, the long-run supply curve will slope upwards. In this case, the existence of external diseconomies will accentuate the upward slope of the supply curve. External economies, however, will offiset to some extent, and perhaps, if strong enough, even revers, the upward slope of the supply curve. The long-run supply curve could thus slope either uwpards or downwards or be

horizontal, depending on how strong the external economies are.

To sum up, then, in the short run the supply curve of the perfectly competitive industry will always slope upwards to the right. In the long run assuming that there is free entry, it is likely to slope upwards unless all factors are homogeneous, but it may conceivably slope downwards if there are sufficiently great external economies as the industry expands.

We have now seen how a supply curve is built up. Our next main task is to analyse the relationship between demand curves and supply curves. In particular we shall be able to study with greater precision the differences between short- and long-run supply curves. First, however, it will be useful to say something about elasticity of supply, where supply curves



are real curves and not straight lines as we assumed in our discussion of elasticity of supply in Chapter I.

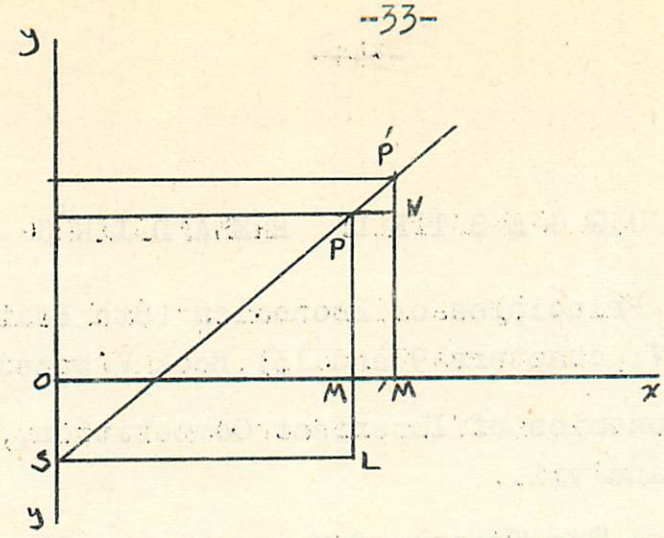
We saw there than any supply curve which is a straight line passing through the origin, whatever its slope, will possess unitary elasticity of supply. We may now add that any straight-line supply curve which cuts the (vertical) y-axis (or whose projection cuts the y-axis) will have an elasticity greater

than one. Any straight-line supply curve which cuts the x-axis (or whose projection cuts the x-axis) will have an elasticity which is less than one. This can be proved with the aid of Figures 59a and 59b.

In Figures 59a and 59b SP' is a supply curve and SL, PN and OM' are all parallel. Now, we saw in Chapter I than elasticity of supply equals.

$$\frac{\text{Increase in amount supplied}}{\text{Amount supplied}} \div \frac{\text{increase in price}}{\text{Price}}.$$

Therefore over the range PP' on the supply curve SP' in Figure 59a, elasticity of supply = $\frac{MM}{OM} \div \frac{NP'}{MP}$. This can, however, be Written alternatively as $\frac{PN}{OM} \div \frac{NP'}{MP} = \frac{PN}{P'N} \times \frac{MP}{SL\ (=OM)}$. But since the triangles SLP and PNP' are similar $\frac{PN}{P'N}$ is equal to $\frac{SL}{PL}$. The formula can therefore be written $\frac{SL}{PL} \times \frac{MP}{SL} = \frac{MP}{PL}$. In Figure 59a MP is greater than PL, so that elasticity is greater than one. It is clear, in fact, that elasticity of supply will always be greater than one, if the supply curve (or its projection) cuts the y-axis. For, whenever this happens, MP will be greater than PL and $\frac{MP}{PL}$ will be greater than one. In some cases, as in Figure 59b, the supply curve cuts the x-axis. In figure 59b the distance PL is now between P and the point L where a perpendicular SL from the y-axis cuts PM projected below the x-axis. The y-axis has been projected below O to meet the supply curve projected below the x-axis. In such cases PL will exceed MP. Elasticity of supply $\frac{MP}{PL}$ will always be less than one, when the supply curve



cuts the x-axis. Finally, whenever the supply curve passes through the origin, PL and MP will always be of equal lengths. Elasticity of supply will always be one.

It is simple enough to apply this analysis to finding the elasticity of supply curves which are true curves. If one Draws a tangent to any point on such a curve, the position at which the tangent cuts an axis will indicate whether elasticity at the point of tangency is greater or less than or equal to one. If the tangent to a supply curve cuts the y-axis, the supply curve has an elasticity greater than one at the point to which the tangent is drawn. If the tangent cuts the x-axis, the supply curve has a point elasticity of less than one. If the tangent passes through the origin, point elasticity is equal to one where the tangent touches the supply curve. This provides a simple rule for discovering whether elasticity of supply is greater than, or less than equal to one. The precise numerical elasticity can, of course, be calculated in the usual way from the formula for elasticity of supply.

SUGGESTED READING

Alfred Marshall, Principles of Economics (8th Edition), London, 1920, Book IV, chapters 9 and 13, Book V passim.

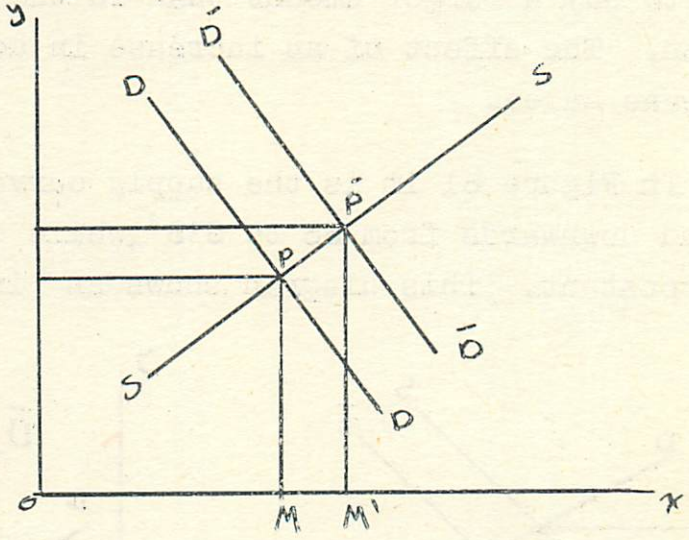
Jon Robinson, Economics of Imperfect Competition, London, 1933, chapters vi and vii.

E. H. Charmberlin, The Theory of Monopolistic Competition (5th Edition), cambridge, Mass.,1946, chapter i.

C H A P T E R V I I
COMPETITIVE EQUILLIBRIUM

1. EQUILIBRIUM AND CHANGE

We have seen how the demand and supply curves for a product made by a perfectly competitive industry are determined. We mut now use these curves to see in greater detail how equilibrium between demand and supply is brought about. We touched on this problem at the end of Chapter I and showed how, with given demand and supply curves, the equilibrium price of a good would be determined by the intersection of these curves. The amount demanded at this price would equal the amount supplied, and the market would be in equilibrium.

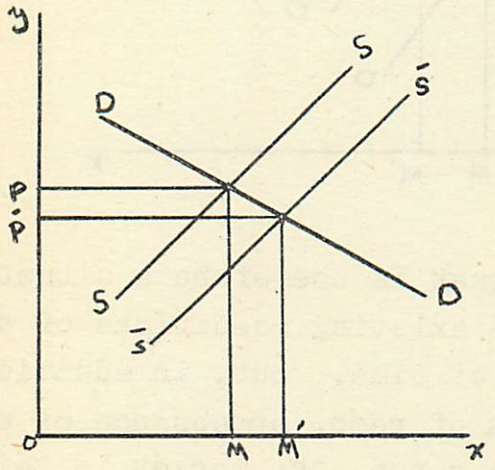


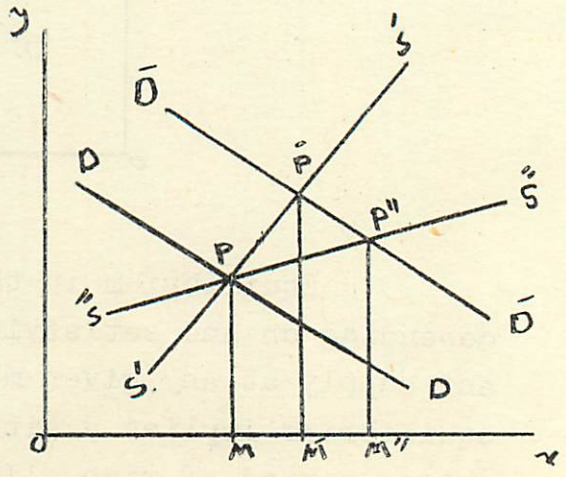
Equilibrium in the market is therefore a situation depending on and satisfying the existing conditions of demand and supply at any given moment of time. But, in addition, equilibrium implies a situation of rest, or absence of chage, over a period of time. In this sense, the market is said to remain in a state of equilibrium only so long as demand and supply conditions, are unchanged. If either demand conditions

or supply conditions, or both of them, alter, the market passes from the old to a new equilibrium.

Such a change can be shown in a diagram. For example, in Figure 60 the demand curve shifts upwards and to the right from DD to D'D', whilst supply conditions remain exactly the same. For the sake of simplicity, it is assumed that the new demand curve is parallel to the old. The new equilibrium price is higher and the new equilibrium amount demanded and supplied is larger than the old. This shows itself in the fact that P' is above and to the right of P. This change can be called for short a rise in demand price or an increase of demand. Consumers are now willing to buy any given amount of the good at a higher price than before. Similarly, at any given price they are now willing to buy a larger amount than in the old equilibrium situation. The effect of an increase in demand is to raise and to increase sales.

In Figure 61 it is the supply curve which shifts to the right and downwards from SS to S'S', while demand conditions remain constant. This diagran shows an 'increased supply' or

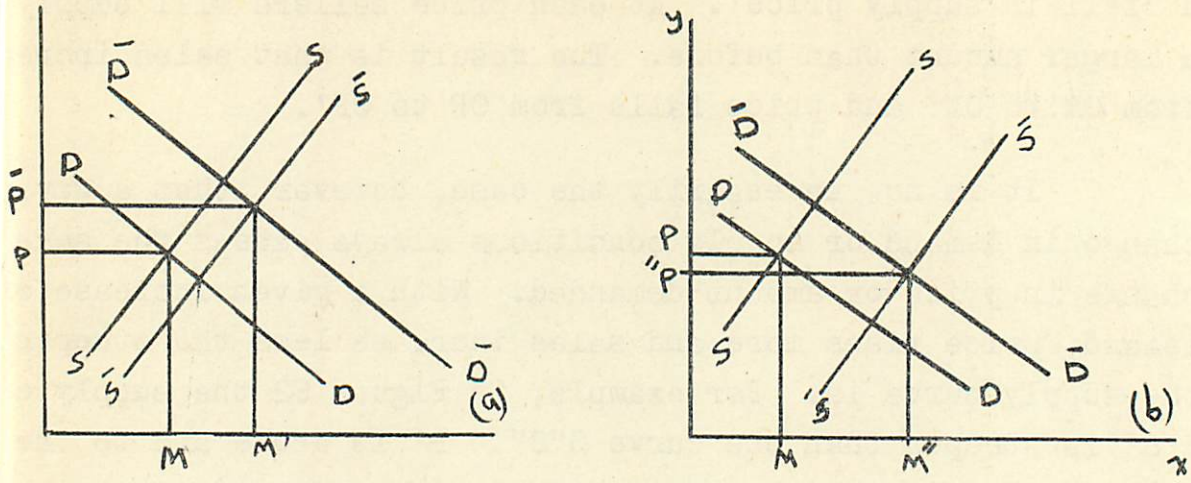




a 'fall in supply price'. At each price sellers will supply a larger amount than before. The result is that sales increase from OM to OM' and price falls from OP to OP'.

It is not necessarily the case, however, that a given change in demand or supply conditions always causes the same change in price or amount demanded. With a given increase of demand, price rises more and sales increase less the steeper the supply curve is. For example, in Figure 62 the supply curve S'S' is steeper than the curve S"S". P' is above and to the left of P", showing that price would rise more and sales increase less with the supply curve S'S' than with the curve S"S", when a given increase in demand took place. Since the slope of both supply curves relates to the same units on the x-axis and y-axis, the steeper curve is less elastic, at each price, than the flatter curve. With inelastic supply, a given increase in demand raises price more than with elastic supply. Similarly, with a given increase of supply, price will fall more with an inelastic demand curve than with an elastic one.

If both demand and supply increase, sales are bound to increase. But price may or may not rise. It will rise if the amount which would now be demanded at the old price exceeds the amount which would now be supplied at that price. But it will fall if the amount which would now be supplied at the old price exceeds the amount which would now be demanded at that price. And if the amount demanded at the old price still equals the amount supplied, price will not change. The first case may be called a smaller increase of supply than of demand, the second case a greater increase of supply than of demand, and the third an equal increase of demand and supply. These first two cases are illustrated in Figure 63, where price rises from OP to OP'

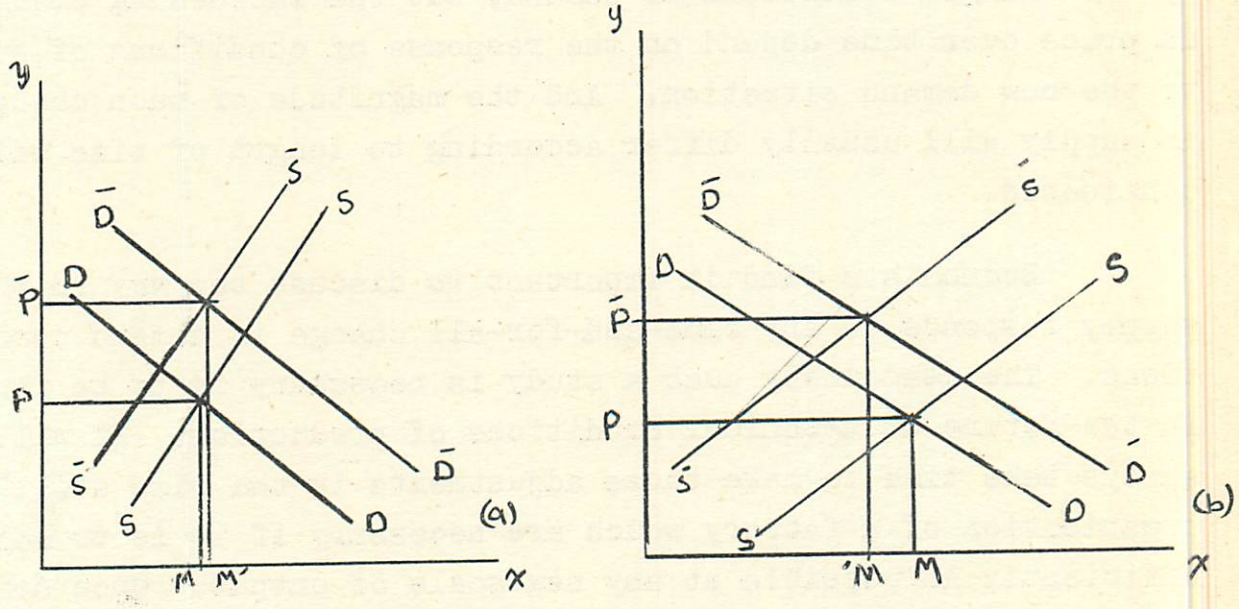


when demand increases more than supply (Figure 63a), but falls from OP to OP" when supply increases more than demand (Figure 63b). The amount of the good exchanged rises from OM to OM' in the first case and from OM to OM" in the second.

In just the same way, an increase of demand with a simultaneous decrease of supply will raise price and increase slaes if the new demand price for the old equilibrium amount is higher than its new supply price. Similarly, sales will diminish and price will rise if the new supply price for the old amount is higher than its new supply price. In the first case the increase of demand is greater than, and in the second case smaller than, the decrease of supply. Both cases are illustrated in Figure 64, where sales increase from OM to OM' in Figure 64a, but diminish from OM to OM' in Figure 64b. In Figures 64a and 64b OM and OP represent the old, and OM' and OP' the new, price and amounts.

It is the function of the market to bring about equilibrium at every moment of time between the forces of demand and supply at that moment. But since one or other of these

forces is continually changing, at any rate in respect of some goods, one or more market prices will be changing at every moment. These price changes may be due to changes in consumers' tastes or incomes on the demand side, or to changes in technical knowledge, or the relative scarcity of factors of production on the supply side. All such changes can be represented in the way we have shown by shifts of the demand and the supply curves respecively. such changes may take the form of a gradual increase of demand or supply, or a grandual diminution of them, when they will be accompanied by a corresponding gradual change in price. Again, they may take the form of alternating increase and diminution of demand or supply. In this latter case, market



price will fluctuate through time. Price changes of these types can be adequately explained by the kind of analysis we have discussed in this section.

2. EQUILIBRIUM AND TIME

There is one other type of change in demand and supply conditions which is important enough to merit special study. We have already seen that supply conditions will be different the longer the period of time which one takes into consideration. For example, if demand studdenly increases, price is likely to rise sharply in the short run because firms will be expanding output along fairly steep short-run marginal cost curves. But in the long run firms will be reorganised so as to produce the new and higher output more efficiently. For firms will now be producing along their rather flatter long-run marginal cost curves. In addition, new firms will have been able to enter the industry. The initial change in equilibrium prices is caused by the changed conditions of demand, but the succeeding changes in price over time depend on the response of conditions of supply to the new demand situation. And the magnitude of such changes in supply will usually differ according to length of time being considered.

Economists find it important to discuss the way in which supply responds to any once-and-for-all change in demand conditions. The reason why such a study is necessary is to be found in the nature of technical conditions of production. It will always take time to make those adjustments in the size and the organisation of a factory which are necessary if it is to work as efficiently as possible at any new scale of output. Once demand has increased (or decreased) a series of adjustments will be put in motion in every firm in the industry, and these adjustments will bear fruit one after the other as time passes.

Now, it will usually be found that economists carrying outthis type of analysis concentrate their attention on the

response of supply through time to a sudden and permanent change in demand conditions. There is an important reason for this. As we have seen, the reason why changes in supply conditions can only be made after a period of time has elapsed, is that technical factors prohibit instantaneous adjustment to changed demand conditions. But there is nothing on the demand side to correspond to this slow process of adjustment on the suplly side. There is no reason why, if supply conditions change, demand conditions should change as well, or, if they do, why they should change differently in the short run and the long run. Changes in consumers' tastes are not dependent on technology in the way that supply conditions are. Admittedly consumers' tastes may, and probably will, change as time goes on. But this will be a change of data and not a change induced by changed supply conditions. There is no necessary reason why the long-run demand curve should differ from the short-run demand curve, however odd the behaviour of supply has been.

This asymmetry between demand and supply is important because it explains the apparent preoccupation of economists with adjustments of supply over time once-and-for-all changes in demand and not vice versa. It does not mean that either demand or supply is more important as a factor determining price. Both are equally improtant. But demand and supply do not respond either equally,or equally quickly, to changed conditions. We must expect that the longer is the period during which demand and supply are coming into equilibrium, the more changes will have time to take place. If we were to study the changes in demand and supply which would take place, in response to any change of data, during many successive very short periods of time, we should find that we had introduced unnecessary and intolerable detail into the analysis.

To avoid this problem, it is now usual to follow the method of analysis used by Marshall. He found that all the most important problems raised by the introduction of time into economic analysis of this kind could be dealt with by considering the way in which, and the extent to which, equilibrium between demand and supply was brought about in three periods of time : the market period, the short period, and the long period. The properties of these three types of time period can be summarised as follows :

(a) The Market Period

Marshall conceived of the market period as being only a single day, or a very few days. The fundamental feature of the market period is that it will be so short that supplies of the commodity in question will be limited to existing stocks, or at most, to supplies which are 'in sight'. On the basis of this 'market' supply and of the existing 'market' demand for the commodity, a temporary equilibrium will be brought about between demand and supply. With perfect competition between buyers and between sellers, an equilibrium price will be determined. This 'market' equilibrium price will, however, only exist temporarily and it will not be reasonable to expect that the market equilibrium price will be similar to prices in preceding or succeeding 'market' periods'.

(b) The Short Period

The short period, Marshall defined as a period which was long enough for supplies of a commodity to be altered by increases or decreases in current output, but not long enough for the fixed equipment producing this output to be adapted to produce larger or smaller outputs. In the terminology of the theory of the firm, we may say that output can be altered in the short period

only by producing at a different point on a given short-run marginal cost curve. The size and kind of the individual plant would be taken as unalterable. No change in the scale of operations would be possible, nor could new firms enter the industry, in the short run.

Equilibrium brought about in this period will represent a 'short-period equilibrium' and supply will be able to adapt itself, at any rate partially, to a change in demand. It can therefore be expected that in the short period, a short-period equilibrium price will be brought about Variations or oscillations in market price caused by conditions specialt to each separate market period will still occur, but these will tend to range round a fairly well-defined position of short-run equilibrium.

(c) The Long Period

In the long period, as defined by Marshall, there is time for firms' fixed equipment to be altered so that output is capable of adapting itself more fully to changes in demand conditions than it was in the short period. There is time to build new machines and factories, and old ones can be closed down or allowed to fall to pieces. A 'long-period equilibrium' between demand and supply will be brought about and in this equilibrium position supply will be able to adapt itself fully to any change in demand. Market equilibrium will now tend to show oscillations round, and variations from, a farily clearly defined long-period equilibrium position.

For the sake of completeness, it may be added that Marshall also talked of a secular[1] or very long period, during

---

1) Derived from Latin Saecula and meaning age-long.

which not only might all the changes which occur in the odinary long period take place, but all the underlying economic factors, such as the size of population, supplies of raw materials, general conditions of capital supply and the like, would have time to alter. We shall not consider the secular period in any detail, since it is much too long to provide any realy satisfactory generalisation for economic theory. In any case, the technique of analysis needed to studey 'secular equilibrium' is the same as for ordinary long-period equilibrium.

3. 'NORMAL' PRICE

We must now introduce the concept of 'normal' price. Marshall defines 'normal action' as that which one expects to be taken by a person or group of persons under given conditions.[1] Similarly, 'normal results' are defined as those results which may reasonably be expected as the outcome of a given situation. 'Normal prices' are therefore those prices which may reasonably be expected in given conditions of demand and supply. But time is very important here. A different price will be 'normal' in the long period from that which is 'normal' in the short period. In our discussion of equilibrium prices in short and long-period we shall really be concerned with short-and long-period 'normal' prices-those prices which one may reasonable expect to occur in the short and long runs respectively.

An important reason for introducing Marshall's periods of time is that it is thereby possible to throw light on the disputes which have often raged between economists on whether demand or supply is more important in determining price. Marshall likens this dispute to an argument about whether the upper or lower blade of a pair of scissors really cuts a piece of paper.[2] He says that

---

1) Principles of Economics, p. 34.

2) Ibid. p. 348.

if the upper blade is held still and the lower blade moved, it is reasonable, provided one does not wish to be completely accurate, to say that the lower blade does the cuting. Alternatively, if the lower blade is fixed and the upper moved, it is reasonable, provided again that one makes no claim to scientific accuracy, to say that the upper blade is cuting the paper. Similarly in certain circumstances, it is possible to say, with some justification, that either demand or supply determines price.

For example, in a fish market where there is a given stock of fish on hand, and where, owing to the danger that it will go bad if kept for long, that stock must be quickly disposed of, it is not unreasonable to say that price is governed by demand. Since supply is fixed, it is demand which determines price. For the sake of brevity, one may use this simplified argument, but one cannot claim strict scientific accuracy. At the other extreme, in the long run when supply and demand changes have worked themselves out fully, the 'normal' long-period price will be the long-run money cost of production (including normal profits). Ignoring the oscillations of market price, the long-period 'normal price' around which such oscillations will occur can be roughly described as determined by supply-by costs of production. Again this is not strictly accurate, but has some truth in it.

The only realy accurate answer to the question whether it is supply or demand which determines price is that it is both. At times it will seem that one is more important than the other, for one will be active and the other passive. For example, if demand remains constant but supply conditions vary, it is demand which is passive and supply active. But neither is more or less important than the other in determining price. It is, of course, important ro realise that, in practice, a long-period normal

price will never be arrived at. There will usually be a change in some of the conditions underlying the long-period equilibrium before it has had time to come int being. The long run-like tomorrow-never comes.

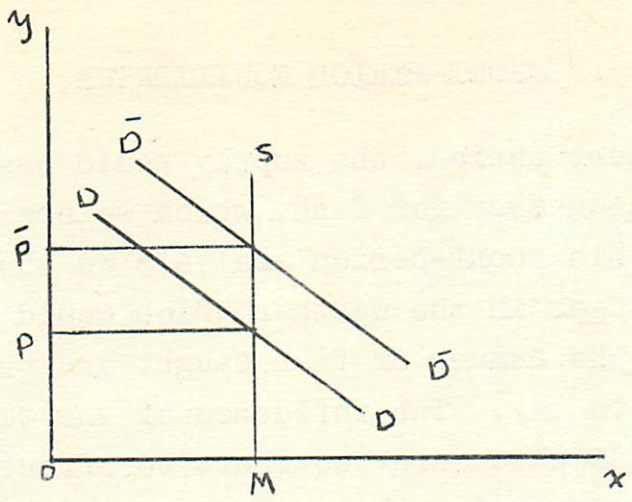
We may now discuss the way in which prices are determined in these various time periods. We shall use the usual method of economic theory and proceed step by step. We shall concentrate our attention on those things which seem most important, by assuming that all other things, except the ones in which we are interested, are equal. As Marshall put it, 'He (the economist) segregates those disturbing causes, whose wanderings happen to be inconvenient, for the time in a pound called Ceteris Paribus.'[1] It should always be remembered that whilst in this way it is possible to construct a simple theory, it will not be a fully accurate theory. The more simple a theory is, the less accurate it must inevitably be. As Marshall says, 'A man is likely to be a better economist if he trusts to his common sense and practical instincts, than if he professes to study the theory of value and is resolved to find it easy'.[2]

4. MARKET EQUILIBRIUM

In order to explain the importance of Marshall's time periods for price theory, we shall study the way in which the supply of a commodity adapts itself throught time to a once-and-forall change in the demand for it.

Let us imagine a fish market with a given supply of fish. We assume that the fish cannot be kept overnight and that the whole supply must therefor be sold quickly, since none of it can be kept back. Competition between buyers and sellers will

1) Op. cit. p. 366.
2) Op. cit. p. 368.



means that the equilibrium price is established where demand is equal to supply. This position is whown in Figure 65 where the supply curve is the vertical straight line MS. This implies that sellers are determined to sell the whole supply of fish, OM, whatever its price-that sellers have no demand for their own fish. This is the assumption which we are making throughout Part One, and it seems a fairly reasonable one. If the sellers did demand their own fish, the supply curve would slope upwards to the right in the normal fashion.

In this market, with demand conditions whown in the curve DD, the equilibrium price of OP will be reached, and the whole supply of fish will be disposed of. If, however, there is a sudden and permanent rise in the demand for fish to D'D', perhaps because of the sudden advent of a serious cattle disease, the price of fish would rise tc OP'. There would be no increase in the supply of fish, since the only available supplies are already in existence and cannot be increased except after a day or two. The whole supply of fish would still be sold, but the price would rise. Owing to the fact that supply is fixed, demand exerts its full influence on the price of fish, which rises considerably.

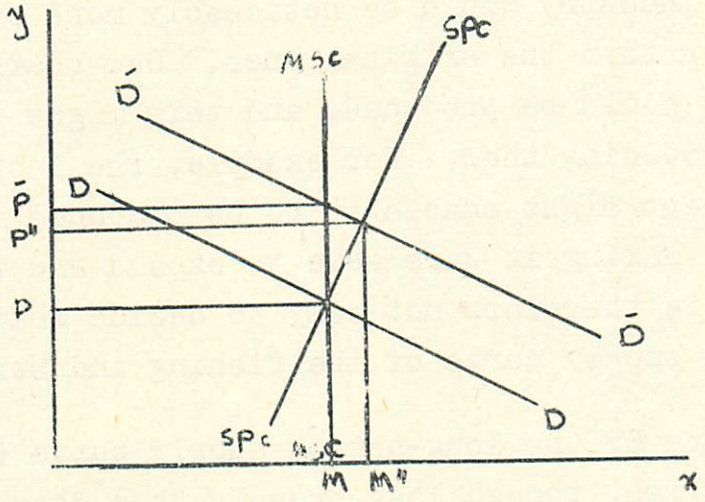
5. SHORT-PERIOD EQUILIBRIUM

In the short period, the supply would begin to react to the change in the demand for fish, which we are assuming to be permanent. In this short-period analysis we shall ignore factors like changes in the weather which would cause daily fluctuations in the number of fish caught and therefore affect prices from day to day. The influence of the weather can only be seen in periods which are too short to affect our analysis at this stage. Similarly, we shall ignore changes which can only occur over very long periods of time. We shall, for instance, assume that there is not enough time for low wages in the fishing industry to persuade sailors to become bank clerks or farmers instead. Such influences affecting the supply of sailors can only enter our analysis if it relates to long periods of time. We therefore abstract from these problems by 'impounding them' for the moment in ceteris paribus.

The problem which has to be faced in the short run is, 'How can we attract sailors into the fishing industry until more of them have been trained? And how can we renovate, or otherwise put into service, more boats?' The answer will usually be, 'By offering wages to sailors which are high enough to attract them from jobs on passenger and cargo boats and by ensuring that the owners of old and not very efficient boats earn enough to make it worth while using them'. By definition, no new firms can enter the fishing industry in the short run but those already in it can use their existing equipment more intensively.

The price which will slove this short-run problem, the 'short-period normal price' of fish, will be that price for fish which will rapidly ensure that enough sailors and boats

go out fishing for all of them to earn the minimum amount of money they are prepared to accept for an average day's work. The market price will still oscillate round this 'normal' price, but the normal price itself will be high enough to ensure that the additional factors of production persuaded to work in the fishing industry find it worth remaining there. This short-run 'normal' price must inevitably be higher than the original market price, for sailors can only be attracted away from other ships



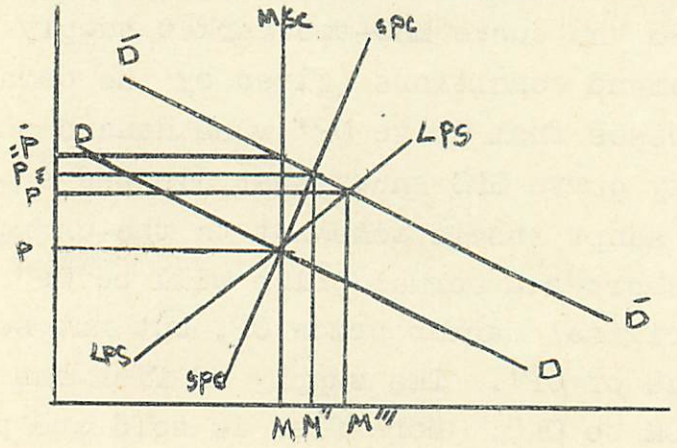
by higher wages, and old boats can only be enticed back to sea by higher returns.

In figure 66 the curve MSC-the market supply curve-shows that in the new demand conditions (given by the demand curve D'D') the market price rises from OP to OP' when demand changes. The short-period supply curve SPS shows that, in the short period, supply is able to adapt itself somewhat to the changed demand conditions. The short-run normal price will be OP" which is higher than the original market price OP, but not so high as the second market price of OP'. The supply of fish has also increased slightly from OM to OM"; More fish is sold and price is not quite so high as in the market period.

6. LONG-PERIOD EQUILIBRIUM

In the long period, supply conditions are fully adapted to meet the new demand conditions. The problems and changes which have to be considered in the long period are rather different from those which arise in the short period. For example, new fishermen would have to be attracted from other jobs, and new firms set up. But it is not likely that the new workers in the industry would be noticeably more or less competent fishermen than the original ones. New boats, nets and other equipment would be produced, and this might well affect the industry producing them. For example, the increased demand for such equipment might enable it to be produced more cheaply if the industry making it were able to expand and to reap external economies. It is therefore not easy to decide what the shape of the long-period supply curve of the fishing industry will be.

In Figure 67 the long-period supply curve (LPS) slopes upwards to the righ, though less steeply than does the short-period curve shown in Figures 66 and 67. The long-period 'normal' price is OP"', as compared with the short-run price of OP", which is higher, and the originalmarket price OP, which



is lower. In the long run it is possible to obtain fish more cheaply and in larger quantities than in the short run, but only at a rather higher price than in the original situation. In the long run, therefore, supply is able to adapt itself as fully as possible to the changed conditions of demand. The extent to which the long-run price differs from the original price thus depends on the ease with which additional supplies can be acquired in the long run. Market prices will still oscillate and fluctuate but they will do so in the long run round a long-period 'normal' price.

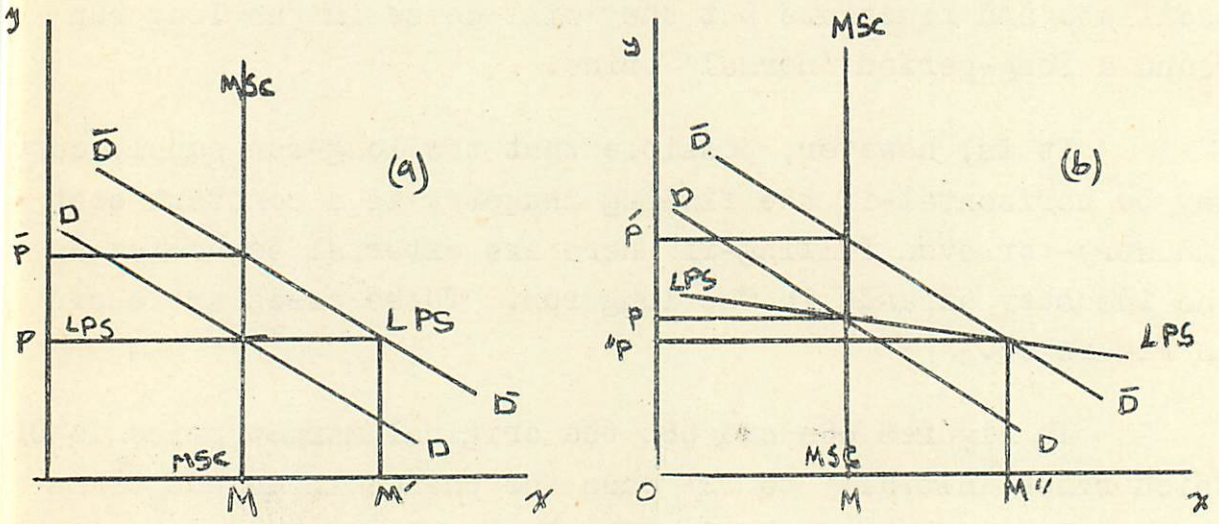
It is, however, possible that the long-run supply curve may be horizontal-if the fishing industry is a constant cost industry- or even falling-if there are external economies as the industry expands in the long run. These cases are shown in Figure 68.

In Figures 68a and 68b the original market price is OP, which rises initially to OP' when the change in demand takes place. The market supply is OM. In the long run, price falls again to OP, since the industry has a horizontal supply curve (LPS). The amount OM' is now sold. With constant long-run costs, the long-run normal price is therefore the same as the original price. In Figure 68b the long-run normal price is OP", which is lower than the original price because of external economies. The amount sold is OM".

It follows, therefore, that the long-run normal price can be higher than, lower than, or the same as, the original market price, depending on supply conditions in the industry in question. The short-period normal price, however, will invariably be higher than the long-period normal price. It is higly improbable that there will be any change in the organisation of the industry, which will increase the efficiency of

the industry, which could be made in the short period but could not be made in the long period.

It is probably worth pointing out that 'normal' prices are not the same thing as 'average' prices unless prices are constant. Normal pirces are those prices to which one may expect actual prices to tend; average prices will be an arithmetical average of all actual prices. They will not only be



influenced by fortutous fluctuztions and oscillations, but will also take account of the general trend towards the 'normal' price.

This analysis by means of time periods has enabled us to show, in an elementary way, how prices vary over time. It has enabled us to drop the assumption, made practically throught the previous chapters, that demand and supply conditions are given. It has allowed us to assume that both can, and often will, change. In technical economic language, we have proceeded from a static analysis, where demand and supply conditions are given, to an analysis of comparative statics, where demand and supply conditions are allowed to change and the new equilibrium is compared with the old.

The construction of a truly dynamic theory of economics, where more continuous changes in demand and supply conditions, like those which occur in the real world, are analysed, is the ultimate goal of most theories of economics. But such theories are usually very complicated, and mean that a great deal of algebra is needed. We shall find it simpler in this book to confine the analysis of the problems of time to a discussion which uses the more easily manageable concept of Marshall's three different periods to study only comparative statics. After all, these give us the basis for quite a useful analysis.

SUGGESTED READING

Alfred Marshall, Principles of Economics (8th Edition), London, 1920, Book V, chapters 5-13.

Ragnar Frisch, 'Alfred Marshall's Theory of Value', Quarterly Journal of Economics, November 1950, p. 495.

CHAPTER VIII
MONOPOLY

1. THE ASSUMPTIONS

HAVING seen how pricesand output are determined in a perfectly competitive industry, we must now see how they are arrived at when competition is not perfect. As we saw in Chapter V, the two limiting types of market situation, those of pure competition and pure monopoly, are not likely to be found very often in practice. In the real world it is the region of 'imperfect' competition lying between these limits in which one is interested. We turn now, therefore, to a study of the way in which prices and output are determined in conditions of imperfect competition. In making this transition from perfect to imperfect competition, we shall find it useful to keep some of the assumptions we have made so far, but others will be dropped. We shall now say explicitly which of the assumptions will be retained in this discussion of imperfect competition.

We shall continue to assume that there is perfect competition between buyers. We shall assume that there are many consumers buying each product, so that no one consumer is sufficiently important in the market to be able to have any influence on the price of the product by his own actions. So far as the individual consumer is conerned, the prices of all goods must be taken as given. We shall also continue to assume that each consumer is 'rational', that he bases his purchases on a scale of preferences which enables him to draw up indifference maps, and that from the indifference maps of such consumers it is possible to derive demand curves showing how much of each commodity they will buy at various prices. These individual demand curves can then be added together to give a market

demand curve. The demand curve for the product of an individual industry therefore means exctly the same thing under imperfect competition as it does under perfect competition.

So far as the individual firm is concerned we shall continue to assume that the sole aim of its entrepreneur is to earn maximum profits. In other words, we keep the fundamental assumption of 'economic rationaliy'. Therefore, when diagrams are drawn showing the equilibrium position of the firm, such an equilibrium will always occur when marginal revenue equals marginal cost, whether competition is perfect of imperfect. There is, however, one important difference between the equilibrium position of the individual firm in perfect competition and its equilibrium position in imperfect competition. When competition is imperfect, there are no longer sufficient firms in the industry for a change in the output of any one fo them to have a negligible effect on the output of the industry as a whole. thus, whilst it is true that the nature of the demand curve for the product of the industry as a whole is not affected by the fact that competition is now imperfect, it will be found that the average revenue curve of the individual firm does have a different shape when competition is not perfect. We can no longer assume that the average revenue curve of the firm is a horizontal straight line when competition is imperfect. In conditions of perfect competition, where each individual firm has to take the price of the industry's product, once given, as fixed and unalterable, the firm's average revenue curve is a horizontal straight line. In imperfect competition this is no longer the case.

As has been stated earlier, when there is imperfect competition the average revenue curve of the individual firm slopes downwards throughout its length. This implies an

important corollary. In perfect competition, since the firm's average revenue curve is a horizontal straight line, marginal revenur and average revenue are always equal. In imperfect competition, however, average revenue is always falling, and therefore marginal revenue falls as well, but even more swiftly. In geometrical terms, the marginal revenue curve always lies below the average revenue curve and normally slopes downwards more steeply. It will also be remembered that when, in perfect competition, the industry as a whole is in 'full' equilibrium, not only does marginal cost equal marginal revenue, but these are themselves equal both to average revenue and to average cost. In imperfect competition this is not the case. In equilibrium marginal revenue still equals marginal cost, but since marginal revenue is less than average revenue, marginal cost is also less than average revenue - than price.

These, then, are the main differences between the assumptions made in analysing perfect and imperefect competition. They are not very great differences. Unfortunately, however, there is no single representative case of imperfect competition as there is of perfect competition. Though the fundamental distinguishing characteristic of imperfect competition is that average revenue curves slope downwards throughout their length, they may well slope downwards at very different rates. Instead of having one single case of imperfect competition to compare with the single case of perfect competition, we can find some instances where the firm's average revenue curve slopes downwards only very gently and where competition is almost perfect, and other cases where that slope is very steep and competition is extrememy imperfect. There is no single case of imperefect competition, but a whole range or series of cases representing progressively more and more imperfect competition.

Since the general term, imperfect competition, covers all situations where there is neither 'pure' competition nor 'pure' monopoly, it is usual to distinguish several separate and diffferent smaller categories within the broad field of imperfect competition. We shall discuss these various narrower types of imperfect competition in turn. We shall begin with a type of market form which has always attracted the attention of economists - that of monopoly. Strictly interpreted, a 'monopolist' is the sole producer of his product, and the distinction between the firm and the industry, both producing the same product, so important in perfect competition, goes. The firm of the monopolist is not only a firm, it is an industry. It is the only firm producing the product in question. The firm, that is to say, takes on the characteristics of the industry, and has an average revenue curve which slopes downwards just as the demand curve for the product of an industry slopes downwards.

2. PURE MONOPOLY

Before embarking on the analysis of monopoly, where, although each firm is the only producer of a given good, this good has to compete indirectly with other goods all of which are competing for the same consumers' incomes, it will be useful to look for a time at the limiting case of 'pure' or 'perfect' monopoly, where competition even in this limited form is competely absent. We have already given a definition of 'pure' monopoly, and have shown that it will occur when the average revenue curve of the firm is a rectangular hyperbola with an elasticity of demand equal to one, and when the monopolist takes the whole of the community's income all the time.[1] It is important to remember,

---

1) See pp. 105-6.

however, that unlike perfect competition, which is a reasonable approximation to reality in certain industries, pure monopoly is merely a theoretical limiting case.

No-one is a complete monopolist because in the end all producers must be competing for the limited resources of consumers. Ultimately all goods are competitive with each other, however imperfectly. This means that the only way to be a perfect monopolist would be to produce all goods. It is therefore useful to distinguish between 'monopoly' where one has competition but not very close competition, and 'pure monopoly' where one has no competition at all. For a real world monopolist is an imperfect competitor rather than the sole and absolute controller of all commodities.

The explanation of the way in which a 'pure' monopolist would fix his output and orice need not detain us long, for the answer is very simple Since a 'pure' monoposlit earns a fixed and constant amount of money (the whole of all consumers' incomes), his profits will be at a maximum when his total costs are as low as possible. This will presumably happen he is producing a very small output (perhaps only one unit) and selling it for an extremely high price. It could never pay a pure monopolist to produce more than a very small output so long as his costs were positive.

One merely has to state the kind of price-output policy a pure monopolist would follow to see how unrealistic the idea of pure monopoly is. No one producer in any country is in such a powerful position that he can sell a minute amount of his product for a fantastically high price taking the whole of consumers' incomes in the process. Pure monopoly is merely a theoretical limit. We must turn for a more realistic analysis to the producer

who is called a 'monopolist' in the real world; to the producer who controls the whole supply of a single commodity which has no close substitutes. The question of how such a producer will fix his price and his output is much more important.

3. MONOPOLY AND PERFECT COMPETITION COMPARED

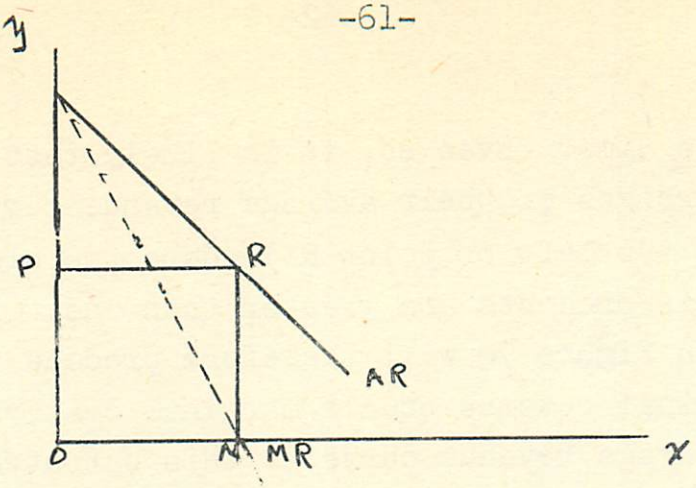
If we define a monopolist as the sole producer of a product which has no closely competing substitutes, it is possible to generalise about his price-output policy and to compare the equilibrium of the firm under monopoly with the equilibrium under perfect competition. It is important to remember at the outset that under monopoly, since this is an extreme form of imperfect competition, the average revenue curve of the firm will slope downwards throughout its length. One can then say first of all that no monoposist will ever fix the output of his product at any level where the elasticity of his average revenue curve is less than one. For if he were to do so, it would always be possible to oncrease his total receipts by restricting output. We have already seen that when elasticity on any average revenue curve is less than one, total receipts will be always falling as output increases-marginal revenue will be negative. It follows that total revenue will rise if output is decreased. So, provided that a monopolist's marginal costs are not negative (which is most unlikely), he will always be able to earn larger profits by reducing his output, if he is producing an output at which the elasticity of demand for his product is less than one. For, since marginal costs are invariably positive, a reduction in output will reduce total costs. At the same time, since elasticity of demand is less than one, a reduction in output will raise total revenue. Profits will therefore rise as output is reduced. Only if a monopolist has negative marginal costs, which seems impossible. will it pay him

to produce an output where the elasticity of demand for his product is less than one.

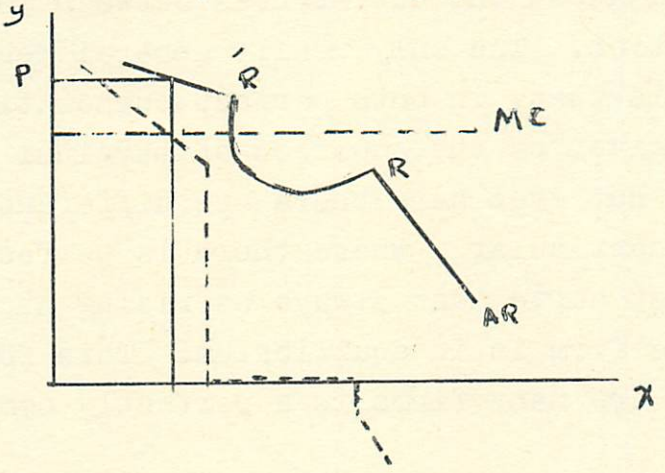
If the elasticity of a monopolist's average revenue curve were equal to one over a small range of outputs, he would be completely indifferent which of those outputs the produced-so long as marginal cost was zero. If marginal cost were positive, a much more probable situation, it would always pay the monopolist to reduce his output until the elasticity of demand for his product became greater than one. For any reduction in output would reduce total costs, but since elasticity is equal to one, it would leave total revenue at exactly the same level as before. We are assuming here that, with all normal aferage revenue curves, a range where elasticity is equal to one will have elasticity greater than one at all points to the left of that range. A monopolist's equilibrium position will always be where the elasticity of demand for his product is greater than one. For only in such conditions will it be possible for a monopolist with positive marginal costs to find an output where, if he decreases production, revenue falls by more than costs fall. If elasticity were not greater than one, a reduction in output would always raise profits.

Figure 69 shows an extreme situation where a monopolist has no costs of production, the time-honoured case of a mineral spring, for example. The equilibrium position will be where the monopolist produces the output OM and sells it at OPs. per unit. This will maximise his receipts. Point R on the average revenue curve will have an elasticity equal to one and total receipts will be at a maximum. Marginal revenue will thus be zero and will equal marginal cost.

Figure 70 shows the more usual situation where the monopolist has positive marginal costs. We have, however,



assumed that these costs are constant, but only in order to simplify the analysis. Here the monopolist is not interested in any part of his average revenue curve beyond point R, because beyond R elasticity of demand is less than one. Between R and R' the curve has an elasticity equal to one and, since his marginal costs are positive, it will pay the monopolist to reduce output at least as far as R'. If the producer's average revenue curve had an elasticity of demand equal to one at all points to the left of R, it would pay him to reduce his output



until he was selling as little as possible (presumable one unit) for an extremely high price. This would not, of course, mean that the producer was a 'pure' monopolist. As we have seen, such a producer must not only have an average revenue curve of unit elasticity he must also take all consumers'

incomes all the time. Even so, it is likely that all monopolists will have ranges of their average revenue curves (in Figure 70 this is to the left of point R') where the elasticities of demand for their products are greater than one. A monopolist like the one in Figure 70 will therefore produce where output is OM and marginal revenue equals marginal cost, and the elasticity of the average revenue curve at this output will be greater than one. Since demand curves with unit elasticity through out are unlikely to occur the real world monopolist (unless he has zero marginal costs) will normally produce where elasticity of demand is greater than one.

As we have seen, the very large range of situations covered bythe collective term, monopoly, means that it is not possible to find any single case of monopoly or to generalise about monopoly in such a way as to allow it to be compared simply with perfect competition. Nevertheless it is possible to say something about the differences between monopoly and perfect competition. The only really general feature which is common to both is that, in both perfect competition and monopoly, equilibrium is given by the equation of marginal revenue with marginal cost. But even here there are differences between the two cases. In particular , where there is perfect competition the marginal cost curve must always be rising at and near the output where the firm is in equilibrium. This follows from the fact that there are many firms in a perfectly competitive industry.

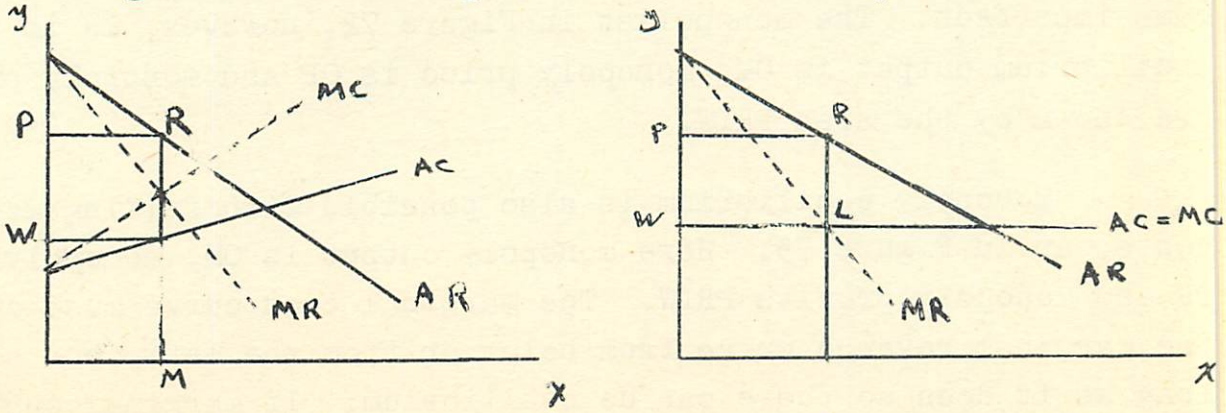
In any perfectly competitive industry the average revenue curve and the marginal revenue curve of the firm are both horizontal. But, as has been shown earlier,[1] the perfectly

---

1) See p. 132.

competitive firm is only in equilibrium when the marginal cost curve cuts the marginal revenue curve from below. If the marginal cost curve cuts the marginal revenue curve from above, profits are not at a maximum. This means that in perfect competition, since the marginal revenue curve is horizontal, if the marginal cost curve is to cut it from below, then marginal cost must be rising at and near the equilibrium output. Falling cost curves are incompatible with equilibrium under perfect competition, for if the firm's marginal cost curve falls continually, it can never cut a horizontal marginal revenue curve from below. There can never be equilibrium. In such conditions the firm will expand until it becomes so large that its average and marginal revenue curves ultimately begin to fall in order to cut the marginal cost curve. The firm will become so large that competition will become imperfect, and the individual firm will be able to affect the price of its product by altering its own output.

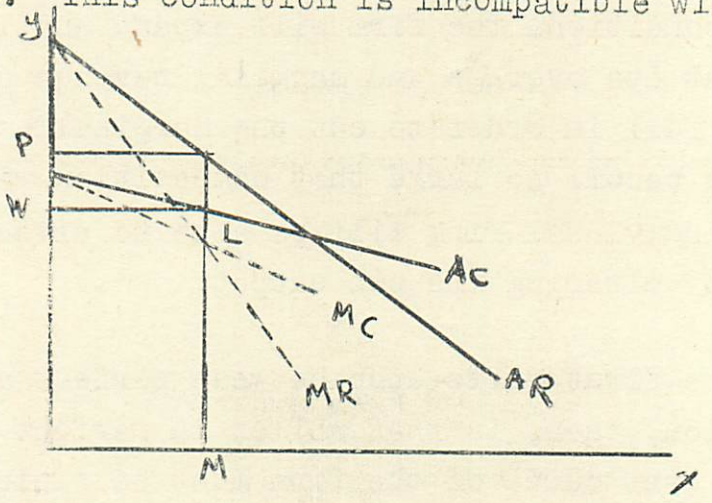
The first difference between perfect and imperfect competition, then, is that whilst in perfect competition the marginal cost curve of the firm must be rising at or near the equilibrium output, in imperfect competition this need not be the case. A monopolistic firm can be in equilibrium with rising, falling or constant marginal costs. The only conditions which



must be satisfied if there is to be equilibrium are that marginal revenue shall equal marginal cost, and that the marginal cost curve shall cut the marginal revenue curve from below. Let us analyse monopoly equilibrium by means of diagrams.

In Figure 71 marginal cost is rising. This ensures that there will be equilibrium in monopoly as in perfect competition. The equilibrium output in Figure 71 is OM and the monopoly price OP. 'Supernormal' or 'monopoly' profits are shown by the rectangle PRLW.

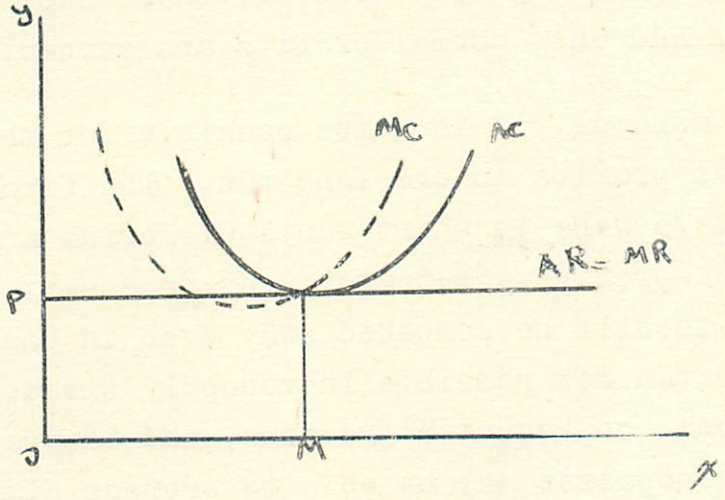
In Figure 72 marginal costs are constant and are equal to average costs. This condition is incompatible with equiibrium



under perfect competition, since, if a firm could afford to produce at all, there would be no limit to its size unless competition became imperfect. The monopolist in Figure 72, however, is in equilibrium output is OM, monopoly price is OP and monopoly profits are shown by the area PRLW.

Monopoly equilibrium is also possible with falling marginal costs, as in Figure 73. Here monopoly output is OM, monopoly price OP and monopoly profits PRLW. The marginal cost curve must cut the marginal revenue curve from below or from the left, but so long as it does so there can be equilibrium. If marginal costs

are falling more rapidly than marginal revenue, equilibrium is clearly impossible. The only situation incompatible with monopoly equilibrium is thus one where marginal costs are falling more swiftly than marginal revenue. If the marginal cost curve is steeper than the marginal revenue curve throughout its length, there can be no output at which the firm is in equilibrium. To sum, up, then, equilibrium for the firm for the firm under perfect competition can only occur when the marginal cost curve of the firm is rising at and near the equilibriuium output. Equilibrium under monopoly can occur whether marginal costs



are rising, falling or constant. The only situation in which monopoly equilibrium is impossible is when the (falling) marginal cost curve is steeper than the marginal revenue curve.

The second major difference between equilibrium under perfect competition and undere monopoly lies in the size of profits. We have seen that in perfect competition the entrepreneur will, in the long run, be unable to earn supernormal profits.[1]

In Figure 74 the perfectly competitive firm is in long-run equilibrium, earning only normal profits. Supernormal profits can be earned in the short run, but in the long run they will be competed away by new entrants into the industry. Thus

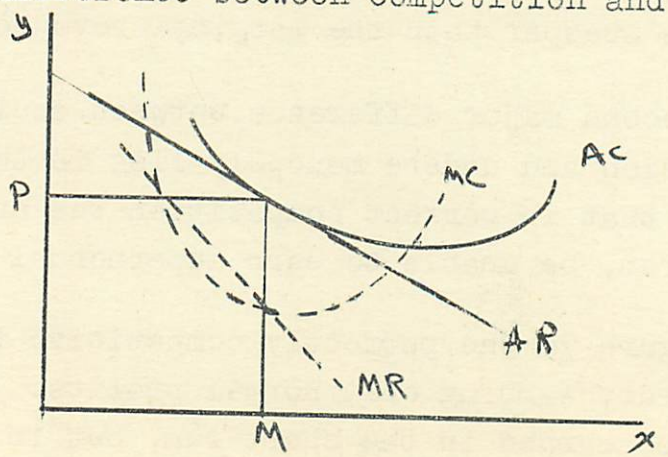
1) See p. 137. We assume homogeneous entrepreneurs.

the long-run conditions of equilibrium for the firm in perfect competition can be summarised as follows:

Total cost = Total revenue.
Average cost = Average revenue.
Marginal cost = Marginal revenue.

In addition, these last four items will be equal to each other and to price. If, in the short run, demand for the industry's product increases, price rises and supernormal profits are earned. In the long run, however, these supernormal profits disappear, and only normal profits are earned.

In monopoly it is quite possible for the firm to earn supernormal profits in the long run. The firms shown in Figures 71, 72 and 73 were in short-run equilibrium and were earning supernormal profits. But it is not necessary to assume that such profits will be competed away even in the long run. Free entry is often not possible in monopoly situations for either institutional or technical reasons, and if free entry is absent then the monopolist may be able to protect himself against competition for a very long time. Of course, if there is freedom of entry and competitive the firm ceases to be a monopolist. The second difference between competition and monopoly is



therefore than monopoly profits can be maintained even in the long run.

This does not mean, however, that a firm cannot, even under monopoly, be earning only normal profits.

In Figure 75 we have the interesting case of a firm producing with a downward-sloping average revenue curve denoting imperfect competition, but earning only normal profits-for its average cost curve is tangent to its average revenue curve. It will be noted, however, that such a firm must be producing at less than optimum size. In other words, it will not be producing at minimum average cost. A downward-sloping average revenue curve can never be tangent to a U-shaped average cost curve at or beyond its lowest point. Thus the second difference between competition and monopoly is that in the long run a competitor can never earn abnormal profits whilst a monopolist can do and usually does, though it is conceivable that he may not.

4. DEGREES OF MONOPOLY

We have seen that in monopoly the firm's average revenue curve will always slope downwards. This means that its marginal revenue curve will always lie below its average revenue curve. So, since the firm is in equilibrium where marginal revenue equals marginal cost, it follows that in equilibrium marginal cost is less than average revenue. In other words, marginal cost is less than price. The size of this difference between marginal cost and price is sometimes used to measure the extent of the firm's monopoly power. In perfect competition, when a firm is in long-run equilibrium, marginal cost equals price. When a firm is in a very monopolistic position marginal cost is considerably

less than price. Economists therefore refer to the difference between marginal cost and price as measuring the degree of monopoly. The larger the difference between marginal cost and price, the greater is the firm's monopoly power.

This gap between marginal cost and price will depend ultimately on the elasticity of demand for the firm's product-on the elasticity of its average revenue curve. This follows from the equations which we gave on p. 98 to show the relationship between marginal and average revenue. The lower the numerical elasticity of demand, the farther does marginal revenue at any output lie below average revenue-the farther, when the firm is in an equilibrium position, does marginal cost fall short of price. Differing elasticities of demand, shown by varying differences between price and marginal cost, are useful measures of the degree of monopoly. The lower is elasticity of demand, the greater is the 'degree of monopoly'.

A second indication of the strength of a monopoly is the size of the supernormal profits the firm is able to earn. In perfect competition firms will be unable for long to earn more than normal profits, because if they do, competitors will enter the industry. With monopoly, new entrants will not normally compete monopoly profits away in this manner. But there will usually be some level of profits at which new firms will find it worth taking the risk of trying to break the monopoly. The stronger the monopolist's position, the greater the profits he will be able to earn without attracting rivals. The size of monopoly profits is thus another (perhaps more reliable) measure of the monopoly power possessed by any firm.

5. DISCRIMINATING MONOPOLY

'Discriminating monopoly' or 'price discrimination' occurs when a monopolist charges different prices for different units of a commodity, even though these units are in fact homogeneous so far as their physical nature is concerned. The extent to which discrimination can occur will obviously depend on the circumstances. In the most favourable situations, it is just conceivable that the price charged could be different for every individual unit of a commodity. Such a situation could be described as 'perfectly discriminating monopoly'. It is more usual, however, to fine that it is different individual consumers who are charged different prices. It would be most unusual to find a single consumer being offered several identical units of a good at different prices. Discrimination between buyers is more usual than discrimination between units of a homogeneous good.

It should be clear that discrimination between customers is incompatible with perfect competition. If there are many sellers of a homogeneous good, it is quite a simple matter for a consumer who feels that he is badly treated by one supplier to go to another one. It would be impossible in perfect competition for some sellers to charge 6d. for an article and others 1s. The peopole who were charged 1s. could easilygo to other producers and get the good for 6d. Price discrimination can only occur when there is imperfect competition. But the relationship is not reversible. Price discrimination does not always occur whenever there is imperfect competition.

It is, of course, possible for a number of imperfectly competing sellers to carry out a jointly agreed policy of price discrimination. In this analysis of discriminating monopoly,

however, we shall assume that it is a single monopolist in whose policy we are interested. The results hold good without serious qualification if discrimination occurs with other types of imperfect competition. We shall assume throughout this analysis that every monopolist always maximises his profits. It is important to decide first in what conditions discrimination is possible at all. We have seen that it is impossible under perfect competition. It may also be impossible even under monopoly. The fundamental condition which must be fulfilled if discrimination is to take place is that there can be no possibility of reasale from one consumer to another. If the same commodity is supplied to Brown at 1s. and to Smith at 6d. and if Brown and Smith can exchange goods freedly, the discrimination will break down. For Smith will buy the good for both of them. So, if price discrimination is to succeed, communication between buyers in different sectors of the monopolist's market must be impossible, or at any rate extremely difficult. In technical language there must be nor 'seepage' between the discriminating monopolist's different markets.

6. CONDITIONS FOR PRICE DISCRIMINATION

It follows that if price discrimination is to succeed there must always be a special reason why consumers in different parts of the monopolist's market cannot communicate with each other. This limits the possibilities of price discrimination fairly narrowly. There are three main types of situation where price discrimination can occur even though there is no fundamental difference between the goods offered to each customer. These are as follows:

(a) Discrimination owing to Consumers' Peculiarities

Discrimination in this type of case can occur for three main reasons:

(i) It can happen where consumer A is unaware that consumer B gets the same good more cheaply. Or, to put it more generally, it can happen when consumers in one part of the market do not know that prices are lower in another.

(ii) It can exist where the consumer has in irrational feeling that though he is paying a higher price he is paying it for a better good. For instance, it is probably irrational to think that one gets a better view of a film from the front row of the 2s. 3d. seats than from the back row of the 1s. 9d. seats.

(iii) Discrimination can occur if price differences are so small that it does not seem worth worrying about them.

(b) Discrimination based on the Nature of the Good

This type of case occurs particularly when the good in question is a direct service. Whilst it is possible for matches to be resold by a customer who is only charged a low price to one who pays a high price, it is impossible to do this with haircuts or 'perms'. Since resale of such direct services is impossible, differences between their prices can exist for different consumers. It is, of course, likely that in some cases the homogeneity of the good is not really complete. For example, it is not entirely realistic to think that a film is quite as enjoyable seen from the fron row of a cinema as it would be from farther back. The service provided to the consumer is not the same.

It may be noticed in passing that in the past the most usual example of this type of discrimination was of the doctor who charged a poor man £1 for removing his tonsils, whilst he charged a rich man £100. The advent of the National Health Service seems to have robbed us in Britain of this practical

illustration. But in any case it has defects. In particular, it is not wholly reasonable to think that doctors are quite as anxious to maximise profits as are many business men. Kindness of heart is probable more usual with doctors than with manufacturers.

(c) Discrimination because of Distances and Frontier Barriers

Discrimination often occurs where consumers are separated by distances or where the fact that a national frontier separates two markets means that tariffs can be levied. In such cases prices can diverage. A good may be sold in one town for 6d. and in another for 1s., and so long as cost of transport is not less than 6d. per unit, resale will not be profitable. Again, where a monopolist serves two different markets, say, a home market with a tariff and a world market without a tariff, he can clearly hope to take take advantage of the tariff barrier and raise his prices in the protected market. Similarly, an import prohibition at home would be very effective in allowing a monopolist to keep up the price in his home market. Such market features are clearly not inherent in the nature of the good sold but are a result of artificial or geographical barriers between one state and another.

It should be noted that all these forms of price discrimination depend ultimately on the monopolist's power to ensure that no-one else sells his product to everyone at a lower price. In most cases monopoly power will depedn on the monopolist's ability to retain his customers, even if rivals enter his market, because they prefer to patronise his fim. In some cases, however, there are other restraints on competition. In the case of doctors, for example, the practice of charging higher prices to rich peopole depends ultimately on everyone in the profession acceptint

this as a convention. In other cases there may be legal sanction for price discrimination-as with the sale of electricity at different rates for lighting and heating. In this case the customer is liable to penalties if he tries to save money by using electricity for lighting instead of heating.

7. THE ANALYSIS OF PRICE DISCRIMINATION

We have now seen in what conditions price discrimination is possible. When will it be profitable? This question can be answered by applying the ordinary theory of the firm to a case where there are two markets (or even more) instead of only one. This complicates the analysis needed but does not affect the fundamentals of that analysis at all. We can still base the theory on the assumption that the monopolist seeks maximum profits and therefore fixes his output so as to equate marginal revenue with marginal cost. Now, however, the firm has two separate markets, so that marginal revenue and marginal cost must be equal in both markets. We shall assume that there are many buyers in each market so that there is perfect competition between buyers within each market. We shall also assume that buyers in the one market are unable to communicate with buyers in the other.

On the sellers' side we shall assume to begin with that the monopolist is a monopolist in both markets. This means that his average revenue curve will slope downwards in each market. We shall also assume that price discrimination is physically possible. It will only be profitable if elasticity of demand in the one market is different from elasticity of demand in the other. It will only pay a monopolist to discriminate between two markets profided that the elasticities of demand are different in the two markets. Let us examine the two possible cases.

(a) Elasticity of Demand the same in each Market

Let us assume that the elasticity of demand in each market is the same at each price as elasticity of demand in the other. This means that the two average revenue curves slope downwards, each changing its elasticity at the same rate as the other. The fact that thetwo markets are unable to communicate with each other does not matter, for the demand in each is of the same quality.

(b) Elasticity of Demand Different each Market

If the elasticities of demand in the two markets at all th relevant prices are different, the monopolist will be able to profit from price discrimination. And provided the wants to earn maximusm profits he will in fact discriminate. It is possible, of course, that the monopolist might be legally forbidden to pursue a policy of price discrimination. Assuming that this is so, he would fix that single monopoly price which would maximise profits by equating marginal cost with marginal revenue for the two markets added together.

We shall assume here, however, that discrimination is not prevented by law. The question which the monopolist now asks is, 'Is elasticity of demand the same in both markets at the single monopoly price?' If elasticity is the same in each market at that price, the monopolist will not discriminate, even thouth the elasticities may be different at all other prices. For if elasticity at the single monopoly price is the same in each market, market, marginal revenues will also be the same. This follows from the formula Marginal Revenue = Average Reveune x $\frac{e - 1}{e}$, where is point elasticity of demand. If average revenue (single monopoly price) is the same in each market and elasticity of demand is also the same, then one can

see immediately from the formula that marginal revenues in the two markets will be equal. This means that if output is transferred from one market to another, there will be no gain in total revenue. What is gained in one market will be lost to the other. There is thus no motive for discirmination here.

Let us now assume that at the single monopoly price the elasticity of demand is different in each market. If elasticity of demand at the single monopoly price is smaller in one market than in the other, discrimination will pay. If the elasticity of demand is very low in one market (market A), the price will be raised even above the single monopoly price. For, since the demand in market A is very inelastic, it is very insensitive changes and a rise in price will not cause much fall in demand. Similarly, if elasticity of demand is very great in the other market (marketB), demand will there be very responsive to price changes. It will therefore pay to lower the price of the good in market B below the single monopoly price. For, since elasticity of demand is very low in market A, a decrease in sales will decrease revenue only very slightly, whilst, in market B, elasticity is high and a reduction in price will add a great deal to revenue. It will pay the monopolist, who is charging the single monopoly price to transfer goods from market A with the inelastic demand to market B with the elastic demand. The loss of revenue from reducing sales by one (marginal) unit in market A which has inelastic demand will be smaller than the gain in revenue from ewpanding sales by one (marginal) unit in the market B which has elastic demand, for prices will rise more sharply in market A than they fall in market B. This follows from the formula cited above. When average revenue is given, marginal revenue will be greater the greater is elasticity of demand. So where average revenue (single monopoly price) is the same in each market, marginal revenue will be greater where elasticity is greater

(market B) and vice versa. But how long will it be worth while continuing this process of transferring units of the good from the market with the low elasticity of demand to the one with the high elasticity? Let us analyse this problem in two stages.

Let us first assume that the monopolist has a given output already produced and merely wants to distribute it in the most profitable way between the two markets. Let us consider a situation where the elasticities of demand in the two markets are different and where marginal revenues in the two markets are therefore different also, but where the monopolist begins by distributing his output between the two markets in such a way that the single monopoly price obtains in each. In order to maximise profits the monopolist will restrict his output in the market with the inelastic demand (market A), and expand it in the market with the elastic demand (market B). He will do this because he is adding more to his revenue in market B than he is taking away from his revenue in market A. Since to begin with, price (single monopoly price) is the same in each market, marginal revenue is greater in market B than in market A. This follows from the formula given above. Thus the discriminating monopolist would move up the marginal revenue curve in market A (by restricting sales and raising price) and down the marginal revenue curve in market B (by expanding sales and lowering price) until the marginal revenues in each market were equal.

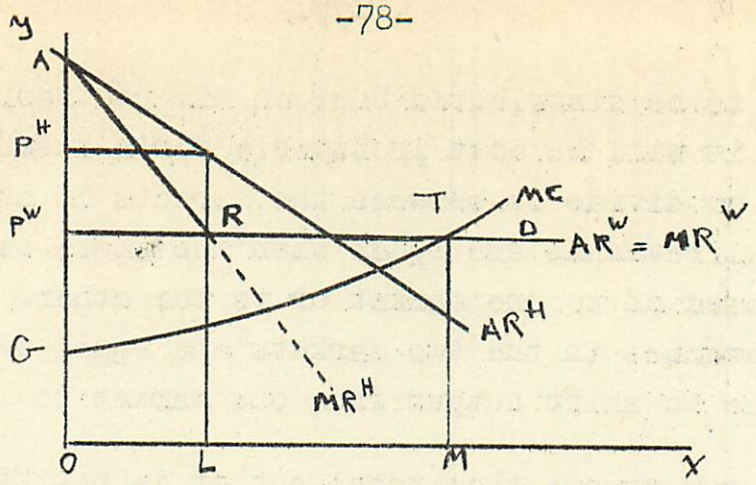
It should be noted that in this equilibrium position, prices will be different in the two markets and the elasticities of demand in the two markets will be unequal also. It follows from our formula that if marginal revenue in each market is to be the same but prices are different, elasticities of demand will be different too. Price will be higher in the market with the less elastic demand and vice versa. We see therefore that if

the output to be distributed between his two markets is fixed in amount, it will be most profitable for a discriminating monopolist to divide it between the markets in such a way that the marginal revenues are equal when the whole output is just being disposed of to one market or to the other. Only if marginal revenues in the two markets are equal will it be unprofitable to shift output from one market to the other.

Let us now assume that total output is not fixed in this way but can be varied. Once again marginal revenues in the two markets must be equal if profits are to be maximised. But it is essential now not only that marginal revenue should be the same in each mardet, but that this marginal revenue should also be equal to the marginal cost of producing the whole output. This is the condition of equilibrium in discriminating monopoly. The position may be shown in a rather complicated diagram, Figure 76.

Figure 76 refers to a producer selling in two different markets. For the sake of simplicity we assume for the moment that he is a monopolist in only one of the markets and that the other is perfectly competitive. In market H, the home market, the producer is a monopolist and the elasticity of demand for his product is not very great, so his average revenue curve AR^H slopes downwards. So does the marginal revenue curve MR^H. In the world market, W, there is perfect competition, and elasticity of demand for the monopolist's product is infinitely great, The average revenue curve AR^W is a horizontal straight line and conincides with the marginal revenue curve for the world market MR^W.

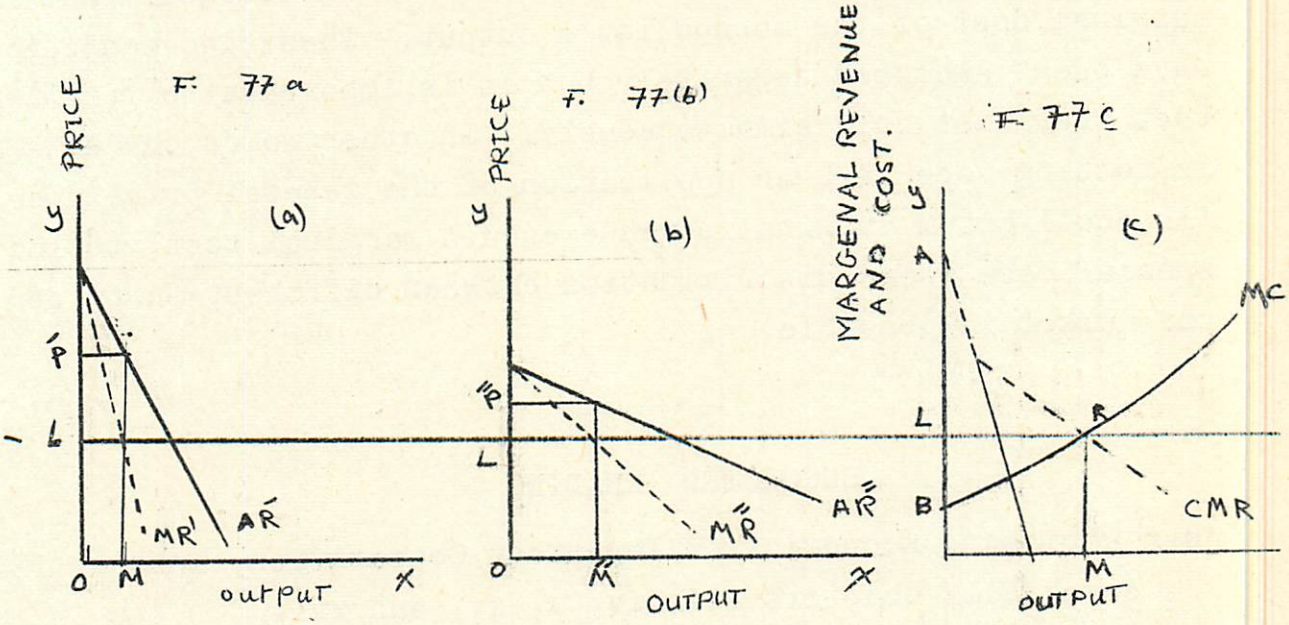
In Figure 76 the marginal cost curve for the monopolist's output is shown by the curve MC. In order to discover how much output it is worth producing altogether, the monopolist must



find where this marginal cost curve intersects the combined marginal revenue curve. Once the size of the total output has been determined, the shares which should go to each market can be decided by allocating the output in such a way that marginal revenues are equal in each market. In Figure 76 the combined marginal revenue curve is given by the composite curve ARTD. On this curve, marginal revenue in the world market is addded on sideways to marginal revenue at home. The intersection of the monopolist's marginal cost curve with the combined marginal revenue curve (at T) gives an output of OM as that total output which maximises profits.

This output OM must now be shared between the two markets. It is clear from Figure 76 that the only way in which the marginal revenues in each market can be equal to each other and to marginal cost MT is for OL to be sold in the home market at a price of OP^H and a marginal revenue of RL. This leaves an amount of LM to be sold in the world market at a price of OP^W and with a marginal revenue MT equal to OP^W. MT is also the same as RL. Price is thus higher (OP^H) in the monopolistic home market than in the competitive world market (OP^W). The monopolist's total profits are equal to the area ARTG. These profits are at a maximum and are contributed to by both markets.

When both markets are monopolistic it is still possible to show, by drawing diagrams, what the monopolist's price-output policy will be, as in Figure 77. Figures 77 a and 77b show the average and marginal revenue curves of the firm in question for two separate markets (markets 1 and 2). These markets have different elasticities of demand at each price. In Figure 77c the profit maximising output is shown to be determined by the intersection of the marginal cost curve for the monopolist's whole output (MC), with the curve showing the combined marginal



revenue earned from the two markets (CMR). The curve CMR is obtained by adding the curves MR' and MR" together sideways. In this equilibrium situation output is OM, and marginal revenue is OL = MR. The output OM has therefore to be distributed between thetwo separate markets in such a way that marginal revenue in each is OL. This means that OM' must be sold in market 1. For with sales at that level marginal revenue in market 1 is OL. Price is OP'. Similarly, OM" must be sold in market 2, at a price of OP". Marginal revenue is OL here too. The monopolist's profit is shown by the area ARB in Figure 77c and is at a maximum. This same solution in terms of marginal and average revenues and costs

can also be shown by combining Figures 77a, 77b and 77c in a single diagram, but naturally this diagram is somewhat complicated.[1]

We have seen that if a discriminating monopolist is to be in equilibrium, two separate conditions have to be fulfilled. First, marginal revenue in both (or all) markets must be the same. Second, these marginal revenues must also equal the marginal cost of the monopolist's output. These two conditions have been analysed separately but it is important to remember that they must hold simultaneously. In other words, MR'=MR''=MC is nothing more than an application of the general principle 'in equilibrium marginal revenue equals marginal cost' to the special case where discrimination between different parts of the market is possible.

SUGGESTED READING

Joan Robinson, Economics of Imperfect Competition, London, 1933, especially chapters iii, iv, v, xv, and xvi.

E. A. G. Robinson, Monopoly, London, 1948, passim.

---

1) Readers will find such a diagram in Joan Robinson, Economics of Imperfect Competition, p. 183.