# THE ECONOMICS OF IMPERFECT COMPETITION

*Books by Joan Robinson*

THE ECONOMICS OF IMPERFECT COMPETITION
ESSAYS IN THE THEORY OF EMPLOYMENT
INTRODUCTION TO THE THEORY OF EMPLOYMENT

# THE ECONOMICS OF IMPERFECT COMPETITION

BY

JOAN ROBINSON

MACMILLAN AND CO., LIMITED
ST. MARTIN'S STREET, LONDON
1942

*First Edition* 1933
*Reprinted* 1934, 1936, 1938, 1942

PRINTED IN GREAT BRITAIN
BY R. & R. CLARK, LIMITED, EDINBURGH

# FOREWORD

BESIDES a restatement of current ideas on economic theory, this book contains some matter which I believe to be new. Of not all the new ideas, however, can I definitely say that "this is my own invention". In particular I have had the constant assistance of Mr. R. F. Kahn. The whole technical apparatus was built up with his aid, and many of the major problems—notably the problems of Price Discrimination and of Exploitation—were solved as much by him as by me. He has also contributed a number of mathematical proofs which I should have been incapable of finding for myself. In general I have endeavoured to build on the foundations laid by Marshall and by Professor Pigou. This is a debt which all economists owe, and which may be taken for granted. I have for the most part referred to their works only where I believe that I have detected them in errors of detail.

Of more recent work, my chief debt is to Mr. Piero Sraffa's article in the *Economic Journal* of December 1926, to Mr. E. A. G. Robinson's *Structure of Competitive Industry*, and to Mr. G. F. Shove's articles in the *Economic Journal* of June 1928 and March 1930. Mr. Sraffa's article must be regarded as the fount from which my work flows, for the chief aim of this book is to attempt to carry out his pregnant suggestion that the whole theory of value should be treated in terms of monopoly analysis. Mr. Robinson's work on the optimum size of firms is the foundation of my treatment of competitive equilibrium, and plays an important part in the Appendix on Increasing and Diminishing Returns. Mr. Shove's articles form the basis of my treatment of rent and of the four cost curves. But a reader who is

acquainted only with those articles would very much underestimate my debt to him, for his teaching in Cambridge for many years past has influenced directly and indirectly the whole approach to many problems of economic analysis. The more specific points that I have derived from Mr. Shove are acknowledged as they occur, but my indebtedness to him must not be regarded as being confined to those points.

A moment has been reached in the development of economic theory when certain definite problems require to be solved, and many writers are at work upon them independently. There are many occasions, therefore, when several explorers are surprised, and somewhat pained, on meeting each other at the Pole. Of such an occasion the history of the "marginal revenue curve" presents a striking example. This piece of apparatus plays a great part in my work, and my book arose out of the attempt to apply it to various problems, but I was not myself one of the many explorers who arrived in rapid succession at this particular Pole. I first learnt of it from Mr. C. H. P. Gifford, of Magdalene College, who was then reading for the Economics Tripos. Shortly afterwards Mr. P. A. Sloan, of Clare College, showed me an unpublished essay in which it occurred. Next it was published by Mr. R. F. Harrod in the *Economic Journal* of June 1930, in an article which must have been written almost simultaneously with Mr. Sloan's paper. In a later article (*Economic Journal*, December 1931) Mr. Harrod set out in an analytical form some of the relations between marginal and average curves which I had discovered by geometry. At this Pole I can claim to have arrived by a route of my own, but his analytical formulation of the fundamental relation between average and marginal value has been of very great service to me since it appeared. Meanwhile a number of explorers were added to the rapidly growing crowd at the Marginal Revenue Pole. Professor T. O. Yntema (who also anticipated Mr. Harrod's formula for the relation of average to marginal value) had,

unknown to me, arrived there long before (*Journal of Political Economy*, December 1928). Dr. E. Schneider, Dr. H. v. Stackelberg, and Professor Mehta, amongst many others, appear to have discovered it independently. Even the naming of this concept presented a minor coincidence. I was dissatisfied with Mr. Gifford's and Mr. Harrod's titles for it, and it was christened for me by Mr. Robinson as "marginal revenue" some time before Professor Viner published an article (in the *Zeitschrift für Nationalökonomie*, September 1931) in which he refers to it by the same name.

The conception of "elasticity of substitution" provides another example of this kind of coincidence, for Mr. J. R. Hicks published his formulation of it in his *Theory of Wages* some time after I had first made use of it. When Mr. Hicks's book appeared my work on the analysis of wages was almost completed, but a study of one of his results led me to remove an error from my argument. In this part of the field my chief debt is to Mr. D. H. Robertson's illuminating article on "Wage Grumbles" (*Economic Fragments*).

A number of writers have recently been evolving methods for dealing with monopoly problems which are at some points similar to my own. Dr. Schneider and Dr. v. Stackelberg have published one or two items which occur in my tool-box. But in my opinion their work is marred by the use of unnecessarily complicated mathematical analysis where simple geometrical methods would serve. I am, however, indebted to Dr. Zeuthen's book on *Problems of Monopoly*. He makes use only of Marshall's "areas" technique, but I discovered at least one important proposition in the course of restating some of his results in "marginal" terms. Professor Chamberlin's *Theory of Monopolistic Competition* provides a plentiful crop of coincidences, but it appeared too late for me to notice them in detail.

There are probably other explorers in the field with whose progress I am unacquainted. If my results are anywhere found to be the same as those of some other writer to whom no

reference is made, it must be understood either that my fellow-explorer is unknown to me, or that his work was published when mine was already completed. But wherever possible I have mentioned the names of the explorers whom I found already at the Pole when I arrived there.

Mr. C. W. Guillebaud was kind enough to read my manuscript and made many helpful suggestions. Finally, in addition to his constructive suggestions, I have had the benefit of Mr. Kahn's criticism at every stage of the work from its inception. For this the reader, as well as myself, must be grateful to him, for he has weeded innumerable errors from my pages.

Some passages in Chapter 7 are taken (with a few alterations) from an article in the *Economic Journal*, December 1932, and are here used with the permission of the Editor.

JOAN ROBINSON

CAMBRIDGE
*October* 1932

# CONTENTS

## BOOK VI

## MONOPSONY

### CHAPTER 17



### CHAPTER 18



### CHAPTER 19



## BOOK VII

## THE DEMAND FOR A FACTOR OF PRODUCTION

### CHAPTER 20



### CHAPTER 21



### CHAPTER 22



## BOOK VIII

## THE COMPARISON OF MONOPOLY AND COMPETITIVE DEMAND FOR LABOUR

### CHAPTER 23



### CHAPTER 24

## BOOK IX

## EXPLOITATION

### CHAPTER 25



### CHAPTER 26



## BOOK X

## A WORLD OF MONOPOLIES

### CHAPTER 27



## APPENDIX

# APPENDIX

## INCREASING AND DIMINISHING RETURNS

### 1

In the foregoing analysis we have made use of the supply curves of particular commodities and of the supply curves of factors of production to particular industries. But these conceptions involve some fundamental questions which we have not discussed. It is possible to make use of a large part of the technical apparatus set out in this book whatever view on these fundamental questions may be adopted, and the attempt to solve them in the following pages is only a provisional one.

### 2

A rising cost curve of a commodity is sometimes described as *diminishing returns*, and a falling cost curve as *increasing returns*. This leads to confusion.[1] Increasing and diminishing returns are more usefully regarded as general principles which may be brought into operation by influences applying to a factor of production, considered separately. The cost of a commodity is built up of the costs of the productive units employed in making it. A rise or a fall in cost (with increases of output) can only come about because the cost, per unit of product, of some item—labour, land, capital, or enterprise—has increased or diminished. As output increases, some of the factors may be found to fulfil the conditions which bring the Law of Increasing Returns into operation, and some the Law of Diminishing Returns. The net result may be a state of affairs in which all the cost curves distinguished in Chapter 10 are rising, or all falling, or some rising and some falling.

[1] Professor Pigou recommends the use of the phrases "increasing supply price" and "decreasing supply price" on the ground that the word "cost" is ambiguous, since it sometimes occurs that average cost is falling, while marginal cost is rising, or average cost rising while marginal cost is falling (*Economics of Welfare*, p. 217). "Supply price," however, is open to the more fundamental objection that it has no meaning for a single firm. It is impossible to speak of the supply price of a monopolist. The best course appears to be to speak of increasing and decreasing cost, and to specify where necessary which cost curve is in question. In the above passage, average long-period cost is the relevant cost.

It is one purpose of this appendix to argue that for a single industry increasing and diminishing returns can be represented in a perfectly symmetrical manner in terms of the supply curves of the factors of production, drawn up in efficiency units appropriately chosen, and they have been treated in this way in the foregoing chapters. But in their nature increasing and diminishing returns are not symmetrical, and we must now examine how they arise.

## 3

The Law of Diminishing Returns, as it is usually formulated, states that with a fixed amount of any one factor of production[1] successive increases in the amount of other factors will after a point yield a diminishing increment of the product. Looking at the matter from the point of view of cost of production, if one factor is fixed in amount and increased amounts of the other factors are used with it, and if no improvement in the efficiency or reduction in the price of these other factors is introduced by the increase in the amount used, after a point the cost of production per unit of output will rise.

At first sight this law appears so obvious as to require no further explanation, but it is possible to restate it in a manner which throws more light on its real meaning. A moment's reflection will show that what the Law of Diminishing Returns really states is that there is a limit to the extent to which one factor of production can be substituted for another, or, in other words, that the elasticity of substitution between factors is not infinite.[2] If this were not true it would be possible, when one factor of production is fixed in amount and the rest are in perfectly elastic supply, to produce part of the output with the aid of the fixed factor, and then, when the optimum proportion between this and other factors was attained, to substitute some other factor for it and to increase output at constant cost.

Thus the Law of Diminishing Returns entails that the various elements required for the production of any commodity should be divided into groups, each group being a factor of production, in such a way that the elasticity of substitution between one factor and another is less than infinite. The Law of Diminishing Returns

[1] The association of the Law of Diminishing Returns with the factor land only arose because land, from the point of view of society as a whole, is by definition fixed in amount. When we are studying the supply curve of a single commodity, there is no reason to expect that land, rather than any other factor, will be scarce. All that the law tells us is that where there is a scarce factor there will be diminishing returns, and labour, capital, and enterprise are just as much subject to it as land.

[2] Elasticity of substitution is defined on p. 256. But for our present purpose it is more convenient to adopt the equivalent but more fundamental definition: the proportionate change in the ratio of the amounts of the factors divided by the proportionate change in the ratio of their marginal physical productivities.

then follows from the definition of a factor of production, and requires no further proof.

Increasing cost for a particular commodity will arise whenever one of the factors of production, defined in this way, is not in perfectly elastic supply to the industry producing that commodity. In the limiting case the supply of a factor may be perfectly inelastic. Given the elasticity of supply of the scarce factor, the extent to which the cost of the commodity will rise, as output increases, will depend upon the elasticity of substitution. If, in the extreme case, there is no elasticity of substitution, so that the production of the commodity requires constant proportions of the factors, the cost curve of the commodity will rise as steeply as the supply curve of the scarce factor. If the scarce factor is rigidly fixed in amount, the supply of the commodity will be perfectly inelastic, and no increase in its output will be possible.

In more usual cases some substitution will be possible and the proportions of the factors will be altered. The cost curve of the commodity will then rise less steeply than the supply curve of the scarce factor, and some increase in output would be possible even though the scarce factor was rigidly fixed in amount. The rise in the cost of the commodity, as output increases, will be less the greater the elasticity of substitution.[1]

An example will make these propositions clear. Suppose that there is a single site available for building a house. Then, if capital and builders' labour were perfect substitutes for land, an infinitely high sky-scraper could be erected on this site at constant cost, and there would be no Law of Diminishing Returns. At the other extreme, if no substitution was possible, only a bungalow could be built on the site, and no increase in the demand for house-room, however great, could lead to an increase in its output. In any ordinary case the proportions of the factors can be altered, but not without limit, and the construction of house-room on a given site is carried out at increasing cost.

Because the proportions of the factors are usually altered (as output increases) when one of them is scarce, the Law of Diminishing Returns is associated with changes in the proportions of the factors. But it is clear that diminishing returns are not due fundamentally to changes in the proportions of the factors, but to the fact that there is a limit to the extent to which the proportions can change.

## 4

We must now consider the supply curve of a factor of production to an industry. For the moment we will assume that there are no economies of large-scale industry.

[1] See p. 123, note.

If the factor which we are considering is perfectly homogeneous in respect to its efficiency in this industry, there is no difficulty in drawing its supply curve. Each unit of the factor (say an acre, or a man) is like every other from the point of view of this industry, and the elasticity of substitution between one portion of the factor and another is infinite. But the supply of the factor to the industry may be less than perfectly elastic, and its cost to the industry may rise as more of it is employed.[1] Here there is no difficulty.

But one of the commonest reasons why the supply of a factor is less than perfectly elastic to an industry is because the factor is not homogeneous in efficiency from the point of view of that industry. It is then necessary to draw up the supply curve of the factor not in its natural units, acres, men, or money capital, but in efficiency units. This can be done as follows: When a given amount of a factor, say land, is being employed by an industry, take any natural unit of the factor, for instance a certain acre, and imagine it to be replaced by other portions of the factor, everything else remaining the same. When another piece of land, working with the same amount of other factors as this standard acre, yields the same product, its efficiency is equal to that of the standard acre. The original acre, arbitrarily chosen, will thus serve as a standard unit, and other areas of land can be reduced to terms of the standard unit, so that the whole supply of land employed in the industry can be expressed in terms of this standard unit of efficiency. It is convenient to call this unit the *corrected natural unit*. It represents natural units of the factors corrected for their idiosyncrasies.[2] The elasticity of substitution, measured in terms of corrected units, will be perfect between one portion of the factor and another. That is to say, if by chance a certain piece of land or a certain number of workers, representing one corrected unit of the factor, were to demand a higher price than the rest they would be dismissed from the industry or be forced to accept the same price as the rest.

If the factor is homogeneous in regard to its efficiency, the corrected units are the same as the natural units, for instance men, acres, or a given amount of money capital, and no correction is necessary. But

[1] See Chapter 8 for the conditions which may produce this effect.

[2] This method of correction is not perfectly satisfactory. The relative efficiencies of different natural units may alter with the amount of other factors employed. The difference between the efficiency of a rich acre and a stony acre may be smaller when wages are low and a high proportion of labour is employed with a given amount of land than when wages are higher and fewer men are employed per acre. It is impossible to say *a priori* in which direction the difference is likely to lie, and our correction would have to be corrected in each case according to the technical conditions of the industry in question and the costs of other factors. This difficulty appears to be insuperable in some cases, but for most of the uses for which we require the conception of the supply curve of a factor to a single industry it can be overcome (see p. 344, note, below).

even if the factors are not homogeneous, so long as there are no economies of large-scale industry, when each factor is increased by, say, ten per cent. in terms of corrected natural units, physical output will also be increased by ten per cent. That is to say, there are constant physical returns. Of course if the price of one of the factors (in these units) is rising it would not in fact be increased by ten per cent. when the others were increased by ten per cent.; an increase of ten per cent. of the physical output would in fact be produced by increasing this factor by less than ten per cent., and the others by more. But if each *were* increased in the same ratio then output *would* be increased in that ratio. It follows that the marginal physical productivity of every amount of a factor, measured in terms of the corrected units, combined in constant proportions with the other factors (again measured in corrected units), is the same, and depends merely upon the proportions of the factors.

Now supposing there are no economies of large-scale industry, so that constant physical returns obtain, draw up a supply curve in terms of corrected natural units. If the factor is heterogeneous in respect of efficiency, but the difference in efficiency between one natural unit and another is the same in this industry and in a number of other industries, the transfer costs of different units will be in the same ratio as their efficiencies,[1] and the supply curve of the factor in corrected natural units will be perfectly elastic. If the factor is scarce from the point of view of this industry, its price per corrected natural unit will increase as more is employed, and the factor will tend to give rise to increasing cost for the commodity.

## 5

We must now consider economies of large-scale industry, and examine the Law of Increasing Returns. The Law of Increasing Returns differs from the Law of Diminishing Returns in that it cannot be reduced to a tautology. The Law of Diminishing Returns, when the factors of production are defined in a certain way, is merely a matter of logical necessity. But the Law of Increasing Returns is a matter of empirical fact. It may be formulated thus: When an increased amount of any factor of production is devoted to a certain use, it is often the case that improvements in organisation can be introduced which will make natural units of the factor (men, acres, or money capital) more efficient, so that an increase in output does not require a proportionate increase in the physical amount of the factors. This law, or rather tendency, like the Law of Diminishing Returns, may apply equally to all the factors of production, but unlike the Law of Diminishing Returns, it does not apply in every

[1] See p. 112.

case. Sometimes an increase of the factors will lead to improvements in efficiency, and sometimes it will not.

It remains to inquire how increases in efficiency can arise. They arise because the factors of production (in the world as we know it) consist of indivisible units, each of which is not equally well adapted to performing all the tasks required in production. If all the factors of production were finely divisible, like sand, it would be possible to produce the smallest output of any commodity with all the advantages of large-scale industry. But actually the factors consist of men (providing labour and entrepreneurship); money capital, which is finely divisible, like sand, but must be turned into instruments of production each of which, for technical reasons, must be of a certain size; and land, which is usually divisible, but which sometimes, for technical reasons, cannot be divided without limit. It is therefore impossible for an industry to equip itself to produce one unit of a commodity without immediately providing capacity to produce more than one unit.

How does this fact account for a fall in cost of production as output increases? The point can be illustrated as follows: Suppose that there is one indivisible unit of a certain factor of production, and that the rest can be increased by small increments, and at constant prices. Then if the cost of the fixed factor is left out of account, the cost per unit of the product up to a certain point will be constant. At first only a part of the fixed amount of the indivisible factor will be used, and as output increases more of this factor will be brought into use. As soon as the whole of the scarce factor is in use, diminishing returns will set in, and the cost of output in terms of the other factors will rise. But meanwhile, if this indivisible factor has a certain cost which must be incurred whether it is fully utilised or not, the average share of each unit of product in this fixed cost will have been falling. Thus at first the average cost of the whole will be falling until the point is reached at which the increase in the cost of the other factors per unit of output outweighs the reduction in cost per unit of the indivisible factor.

The curve representing the average cost per unit of output of the indivisible factor is a rectangular hyperbola, subtending a rectangle equal in area to the cost of the factor, and falling continuously as output increases. The average cost of the other factors is constant up to the output OS, at which diminishing returns begin, and then rises. The curve of average total cost, which is the sum of these two curves, falls up to the output OT and then rises. The curve of marginal cost will be constant up to OS and then begin to rise, cutting the curve of average total cost at its lowest point, for the output OT. When the rise in cost has reached a certain point it will become profitable to use a second unit of the indivisible factor, and the whole process will begin again.

We are already familiar with this effect, for we have used it in the analysis of cost to the individual firm. The indivisible unit is there the entrepreneur, and the other factors are variable. But the same process is at work wherever there is an indivisible unit of a factor which requires a certain price irrespective of its output—a man, who commands a certain wage, or a machine which has a certain cost—and it is this fact which accounts for the technical economies which a firm can introduce when its output increases, over and above the economy of spreading the fixed cost of the entrepreneur over a larger output.

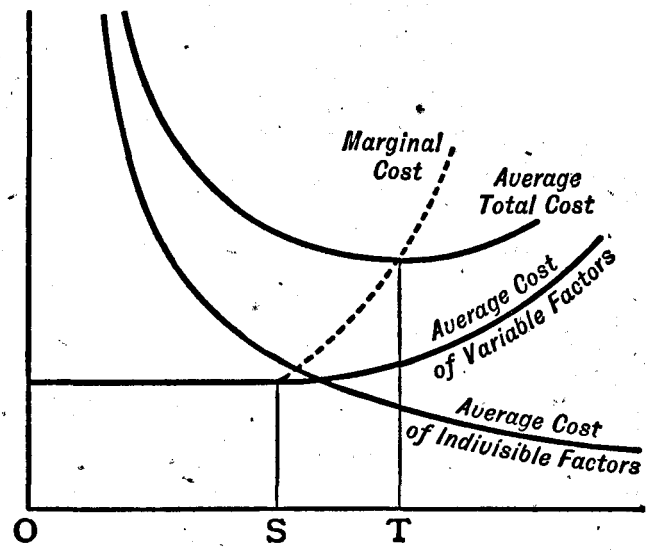


FIG. 82.

The possibility of increasing returns is widened by the fact that various units of the factors are adapted to performing different tasks. Men differ in their natural abilities, and can acquire skill when they concentrate on a single task;[1] acres vary in their natural capacities, and machines can be designed for special tasks. For any kind of production there will be a hierarchy of possible technical methods, each using more highly specialised units of the factors than the last, and production is carried out most efficiently when each separate action in the productive process is performed by a unit of a factor of production specially adapted (by nature, by practice, or by human ingenuity) to that particular task. But since the units of the factors are indivisible, the most specialised method of production will involve the largest outlay, and it is not profitable to make use of the

[1] The increase in efficiency which arises from the fact that "practice makes perfect" is itself a result of the indivisibility of the units of the factors. If labour could be finely divided, like sand, each grain of labour could be occupied constantly at a single task and could acquire the maximum amount of practice.

full equipment of highly specialised factors for a very small output. As output increases a method higher in the hierarchy of specialisation can be adopted, and for this reason cost falls as the output of a commodity increases.

The units of the factors are very often imperfectly specialised, and when output is small a single indivisible unit of a factor, for instance a man, may perform a number of different tasks. The Law of Increasing Returns is often associated with the fact that, as output increases, the number of tasks performed by indivisible units of the factors is reduced. For instance Adam Smith speaks of "the advantage which is gained by saving the time commonly lost in passing from one sort of work to another",[1] and Marshall refers to the waste involved in employing a skilled worker on tasks equally well performed by an unskilled worker, when the output is too small to occupy him constantly at a task which requires his skill.[2] But, fundamentally, the economy of large scale does not arise because particular units of the factors are versatile, but because they are not perfectly versatile.

The maximum rate of decreasing cost would occur if each unit of the factors was completely specialised and capable of performing only one task. If, in Adam Smith's pin factory, each of the workers had been bound by a rigid caste system to a single occupation, then to produce even one pin it would be necessary to employ the whole number of workers—one to draw out the wire, another to straight it, a third to cut it, and so forth. Then, if the wage per man were independent of his output,[3] the total cost of the capacity output of the team of workers would be equal to the cost of one pin, and the maximum possible rate of falling cost would be obtained. When the capacity output of one team was reached a fresh team would have to be employed and there would be no further possibilities of specialisation.

In more usual cases the units of the factors are capable of performing various tasks. Thus small outputs will be less costly than they would be if the maximum possible degree of specialisation had to be introduced at the outset. Each increase in output will require some increase in the amounts of the factors employed, but the increase in output will be more than in proportion to the increase in the factors, because more specialised indivisible units of the factors can be employed as output increases.

[1] *Wealth of Nations*, Book I. chap. i.

[2] *Principles*, pp. 264-65.

[3] The device of paying a unit of a factor according to its output produces the same effect as though the unit were perfectly divisible. If Adam Smith's pin makers were paid at the same rate per pin when each worked separately as when they co-operated the cost of pins would not alter as their output increased.

## 6

We have found that increasing returns are due to the indivisibility of particular units of the factors. In order to account for falling costs for a particular commodity it is therefore necessary to find, at some point in the productive process, a single indivisible unit of some factor of production. So long as there are a number of units of the same kind engaged in any process we know that for the existing output the possibilities of increasing returns are exhausted. There may be some higher degree of specialisation which it is not profitable to introduce unless output is increased. The single unit which will give rise to increasing returns is then, as it were, still over the horizon, but in every case where increasing returns are found, there must be some point in the process of production at which a single unit of a factor is engaged.[1]

It is therefore easy to account for falling costs of production so long as only one firm is engaged upon a particular commodity. The firm may be of less than the size at which average cost is a minimum because some part of its productive equipment, a piece of plant, a salaried employee, or the entrepreneur himself, is capable of co-operating in an increase of output without any increase in cost of that part of the equipment of the firm. When competition is not perfect, firms will be in equilibrium when they are of less than optimum size (if profits are normal), and an increase in the output of a single firm would lead to a fall in average cost.

The question of whether falling cost can occur in a perfectly competitive industry is more complicated. To isolate the effect of increasing returns let us suppose that there is a perfectly elastic supply to the industry in question of all factors measured in corrected units.[2] The industry in equilibrium will be composed of a number of firms, each of optimum size. But the capacity of a single unit of entrepreneurship is limited, and it may be that when the firm is of optimum size there are still technical economies to be gained, in some departments, by a further utilisation of indivisible units of

[1] Cf. Robinson, *Structure of Competitive Industry*, p. 25.

[2] If we say that the supply of the factor is perfectly elastic to a certain industry, we mean that when more labour, capital, land, or enterprise is devoted to one industry, it is attracted by the same payment as before, but once it finds its way into the industry, its efficiency may be increased by specialisation, so that its efficiency price to this industry falls, not because it has become cheaper in general, but because a given portion of it can be turned to better use when a greater total is being employed. When we are studying the question of increasing returns, not in one industry but in industry in general, it is impossible to assume a price for the factor in general, and the inquiry becomes mysterious and difficult in the extreme. As long as we are studying one industry in isolation it can be imagined as drawing upon a general pool of the factor in question, and the cost of a natural unit of the factor can be measured in terms of money price.

the factors or by a higher degree of specialisation of the factors, which are not realised because they are outweighed by diseconomies of large-scale management.[1]

We are tempted to conclude that increasing returns could then occur through the specialisation of firms. Each firm may relieve the strain upon management by abandoning some processes of manufacture to other firms, and so be enabled to carry out the production which it retains upon a larger scale, making use of those indivisible units of the factors which were not fully occupied before. More technical economies can thus be realised, and at the same time it is possible that there will be an additional gain from the fact that individual entrepreneurs, concentrating upon a smaller part of the productive process, may acquire specialised knowledge and skill. But we must examine the matter more closely before we can be satisfied that the specialisation of firms can lead to decreasing cost under the conditions of perfect competition.

The specialisation of firms may be of two types, lateral disintegration[2] and vertical disintegration. Lateral disintegration is the process by which firms, each formerly producing a number of different commodities or types of a commodity, gradually specialise upon a narrower and narrower range of products until (at the last resort) each is only producing a single type of a single commodity. Professor Pigou finds in this process a sufficient explanation for the existence of decreasing supply price, and he quotes as an example the contrast between the British and German cotton industries. The British industry is larger and more highly specialised than the German industry. "The range of work undertaken by the typical factory in Germany is far greater than that undertaken by the typical factory in England. Hence naturally the skill of the operatives is far less in Germany; more time is wasted and factory organisation is less perfect."[3] This principle of lateral disintegration is of the greatest importance in the real world, but will it serve to explain the existence of decreasing costs under conditions of perfect competition? If an industry grows up from the first in a perfect market, we should expect it to develop from the beginning the maximum possible degree of specialisation. If there is anything to be gained by concentrating upon a few counts of cotton we should expect a perfectly competitive spinning industry, while it was still upon a relatively small scale, to consist of a number of firms each producing different counts. It would consist, in short, of a number of industries,

[1] Cf. Robinson, *Structure of Competitive Industry*, chap. vii.

[2] It would be more natural to speak of horizontal disintegration, but "horizontal integration" is commonly used to mean the combination of firms making the same article, and "horizontal disintegration" had best be preserved to mean the contrary process.

[3] Sir Sydney Chapman, quoted by Professor Pigou, *Economics of Welfare*, p. 221.

each in the hands of a monopolist. At first, as the industry grew, there would be some decrease of costs, for as the market in each count increased, the firms could specialise each upon fewer counts; but as soon as the market was large enough to support several firms each producing the same count, the decrease of costs would come to an end, since there would be no further possibility of gaining economies by specialisation. This criticism upon Professor Pigou's argument serves to show one of the absurdities latent in the assumption of a perfect market. It is very unlikely that the saving in cost upon a consignment of yarn, due to extreme specialisation between firms, would be large enough to offset the inconvenience and expense to the purchaser due to ordering each count from a separate producer. If a manufacturer requires a number of different types of yarn at the same time he will prefer to order them all from the same house, unless the prices quoted by a firm which can supply him with all of them are considerably higher than the prices quoted by firms which each specialise upon one or two. But we are now engaged in discussing a perfectly competitive industry, selling in a perfect market. In a perfect market, the customer must be assumed to prefer the goods of the firm that can sell them at the cheapest price, however small the difference in price, and however great the other advantages offered by a firm whose price is slightly greater. Thus in a perfect market the maximum degree of specialisation between firms would come about from the beginning, and the only type of decreasing cost which we should expect to find is that which occurs in a one-firm industry, when the firm is of less than optimum size.

Lateral disintegration, upon our definition of an industry, is the separation of a single industry into a number of parallel industries. Vertical disintegration[1] is the separation of an industry into a series of processes each carried on by separate firms. The cotton industry will again provide an example. In England the cotton industry is divided into sections consisting of firms each devoted to a single process, spinning, weaving, bleaching, dyeing, and so forth. The business of dealing in raw cotton and of selling in foreign markets is also disintegrated and is in the hands of brokers and merchants separated from the producing firms. In Japan, on the other hand, single firms carry out the whole process from buying cotton to selling piece-goods. In the cotton industry the maximum possible degree of vertical disintegration is rapidly attained. When spinning is separated from weaving, neither process can be subdivided any further, but in the case of a complicated object like a motor car, the possibilities of disintegration are almost endless. If a motor firm begins to feel the pressure of diminishing returns from entrepreneurship, as it grows in size, it can abandon the manufacture of some

[1] See Robinson, *Structure of Competitive Industry*, p. 110.

part of the car, the radiator or the body for instance, to a specialist firm, and continue to increase its output of cars without increasing its staff. Meanwhile the specialist firm, as the scale of its output increases, will gain from those technical economies which could not be achieved by the car-producing firms because each individually produced too few of this particular part to allow their full development.[1]

In vertical disintegration as much as in lateral disintegration the degree of specialisation depends upon the size of the market, and again we should expect under perfect competition to find the maximum degree of specialisation at each stage in the growth of the industry. As soon as two or three firms were engaged on each process, we should expect to find that the possibilities of further decreasing cost had come to an end.[2]

Thus we find that when we follow out strictly all the implications of the assumption of perfect competition the grounds for expecting decreasing cost due to the specialisation of firms are very much narrowed, and it is only when there is at some point in the productive process a single indivisible unit of a factor at work (in this case a single specialist firm) that decreasing costs can occur.

If there are no economies from disintegration, either because all the technical economies of large-scale production were already exhausted before the firms grew to the optimum size, or because for technical reasons vertical disintegration is impossible, or because all the economies of disintegration have already been brought about and no further specialisation is possible, then an increase in the output of the industry can only come about by the addition of optimum firms, each like the rest, or of groups of firms carrying out between them the whole productive process.

Even then it is possible that there may be falling costs, for there may be *external economies*. When a new firm enters the industry it may enable all the firms to produce more cheaply so that, while each produces at its minimum average cost, the cost at the minimum is

[1] The vertical disintegration of the British motor industry is continuing every year, and Mr. Ford, whose aim was formerly to control the whole process of manufacture from growing raw rubber for his tyres, has now begun to proclaim its benefits; see *Moving Forward*, pp. 153-54.

[2] We should expect, however, that vertical disintegration would take place less rapidly, as output increases, than lateral disintegration. There will be certain costs of co-ordinating the disintegrated processes which will be reflected in the cost of the commodity. If a commodity is manufactured by a number of processes, each carried on by a different firm, there must be some costs of transport, including the costs of ordering and invoicing, involved in assembling the parts of the finished product. These costs are likely to decline as output increases, for there will be economies in handling goods on a large scale. Thus a degree of disintegration may become profitable for a large output which would not be profitable for a smaller output even though some technical economies could already be gained when the output was small.

reduced. The simplest example of this type of external economies is the case where machinery can be bought more cheaply when the industry presents a larger market to the machine-making industry.[1] But this is properly to be regarded as an example of vertical disintegration. The machine-making industry represents a part of the productive process, already disintegrated from the main industry, which is working under falling costs. We must, then, inquire how the machine-making industry came to have falling costs, and so we pursue the whole inquiry afresh, and find the falling costs to be due either to the existence of a single sub-optimum firm,[2] or to increasing returns due to the progressive specialisation of firms, or to external economies. If they are due to external economies, we must again pursue them until they are finally run to earth.

But there is another type of external economy which does not arise from the scale of a subsidiary industry. If a large labour force is accustomed to work at a certain trade, it may be that a traditional skill is developed, and each individual worker is more competent than he would have been in a smaller industry. Economies of this type, however, which can be found to depend on the size of the industry whose supply curve we are considering, rather than upon the general development of industry, are likely to be rare and unimportant, unless the industry is growing from a very small initial size.

## 7

External economies and the economies of specialisation of firms may be grouped together under the title of economies of large-scale industry, as opposed to the economies of individual expansion, or internal economies, which depend upon the size of the firm. Economies of large-scale industry are likely to have the effect of altering the optimum size of the firm, and the reorganisation of the firm to adapt

[1] In order to study the principle of increasing returns or of diminishing returns in any one particular industry it is necessary to suppose that a change in the amount of any factor employed in this industry has a negligible effect upon the price and efficiency of the factor in general. If this condition is not fulfilled, any change in one industry will alter all costs of production and therefore will have a reaction upon the demand curve for the commodity produced by the industry in question. In practice this condition will often fail to be fulfilled. For instance any increase in the scale of any one industry in a certain district is likely to reduce the costs of all industries in respect of transport, banking, and other facilities enjoyed in common by all the local industries. All the commodities produced in the district will therefore become cheaper, and the demand curve for the commodity produced by the expanding industry will be likely to alter. In such a case it is impossible to treat the demand curve for the commodity as independent of the amount produced. See Sraffa, *Economic Journal*, December 1926.

[2] The fact that the sub-optimum firm must be a monopoly complicates the position. Not every increase in demand will lead to lower prices, though it will lead to lower average cost.

itself to the new optimum size may lead to further economies. These have been described by Mr. Robertson as internal-external economies.[1] They are internal economies, because they depend upon the size of the firm, and external economies because they depend upon the size of the industry. It is easier, *a priori*, to think of reasons why the optimum firm should grow smaller as the result of external economies[2] than of the reasons why it should grow larger. The cheapening of machinery, for instance, will reduce one of the advantages which large firms have over small. If a specialised machine becomes cheaper, the loss due to working it at less than its full capacity becomes smaller, and one of the influences tending towards a large optimum technical size for the firm becomes less strong. On the other hand, any influence tending to reduce the costs of other factors relatively to the cost of entrepreneurship will increase the optimum size of the firm. Professor Pigou, following Marshall,[3] asserts that in general firms tend to grow with the growth of the industry, but the fact that this occurs in the real world can be accounted for by the fact that in an imperfect market the equilibrium size of firms is likely to increase as the industry expands.[4] In the real world there is no reason to expect that firms are at their optimum size, and the fact that firms are growing does not prove that the optimum is becoming larger. Moreover, in the real world inventions have to be taken into account, and a historical movement toward the growth of firms may be due to the introduction of new methods of production suitable to large-scale use. However this may be, the internal-external economies are not likely to be of much importance compared with the economies of large-scale industry which give rise to them.

We may summarise the results of the foregoing analysis as follows. Decreasing costs may occur for the output of a firm of less than optimum size; and for a perfectly competitive industry they may occur when the optimum size of the individual firm is not sufficiently large to allow the full development of all the possible technical economies of large-scale production in every process, so that increasing returns arise from the specialisation of firms, and even when

[1] "Symposium", *Economic Journal*, March 1930, p. 86.

[2] The effect of specialisation upon the size of firms is difficult to discuss, because of the difficulty of defining size. Ordinarily we should measure the size of a firm by its output, but this becomes impossible when the output is changing in nature as the result of specialisation. Measurement by men employed is too crude, and by men *plus* equipment too complicated to be of use. Since the point has not much relevance to the present discussion, it does not seem worth while to attempt to devise an index for the measurement of the size of the firm; cf. Shove, *Economic Journal*, March 1930, p. 115.

[3] *Economics of Welfare*, p. 221; *Principles*, p. 318.

[4] See p. 101. Marshall, who never followed out in the text of the *Principles* the rigid view of perfect competition implicit in his diagrams, may have had this effect in mind.

all the possibilities of specialisation have been exhausted decreasing cost may be due to external economies.

8

In every case increasing returns arise from improvements in productive technique. As output increases the efficiency of the factors can be increased by the fuller utilisation of indivisible units of the factors, or by the adoption of more specialised methods of production. Thus increasing returns are fundamentally different from diminishing returns, which are brought into play, not by a change in the efficiency of the factors, but by an alteration in their price. It is possible, however, to devise a method by which the economies of large-scale industry can be represented in terms of the prices of the factors, so that increasing returns from the point of view of a single industry can be treated in a manner symmetrical with diminishing returns.

We will first consider the simplest type of economies of large-scale industry. Suppose that the same kind of machines are used when the industry expands and the machines become cheaper. Then if we add, say, ten per cent. to the other factors (in terms of corrected natural units) and ten per cent. to the number of machines, we shall have added ten per cent. to output. Thus the machine can be regarded as an efficiency unit of capital, and increasing returns of this simple type could be regarded as arising from a fall in the price of these efficiency units of capital when more are employed.

More complicated types of increasing returns can be treated in the same way, but when the technique of production changes as output is increased it ceases to be possible to see immediately in what the efficiency unit consists. An efficiency unit, however, can be devised as follows: First increase each factor except one by ten per cent. in terms of corrected natural units; now increase the remaining factor, say capital, until ten per cent. is added to the output. If there were no economies it would need an increase of ten per cent. in units of money capital;[1] if there are economies, it will need an increase of less than ten per cent. We shall then say that we have increased capital by ten per cent. in units of efficiency. We are thus provided with an efficiency unit of capital in which to draw up the supply curve of capital to the industry. If it requires less than ten per cent. increase in money value of capital to increase output by ten per cent. (when all other factors are increased ten per cent.), and if the supply of capital in money units is perfectly elastic, the cost of capital will have been increased by less than ten per cent., and the supply price

[1] Since we measure capital in units of money for long-period problems, no correction for non-homogeneity will be needed in this case, and the corrected natural units will be the same as the natural units.

of capital in terms of these efficiency units will be falling.[1] Thus economies of large-scale industry can be represented by a falling supply curve (in efficiency units) of one of the factors to the industry. In the same way, when we were considering the simple case of machines which become cheaper without altering their form, the machine is the efficiency unit, and since the corrected unit of capital

[1] When the technique of production alters as output increases, a difficulty arises similar to that which was discussed in the note to p. 332. The change in efficiency due to a given increase in corrected units of capital (that is to say, money) will depend not only on the amount of capital employed in the first position, but also on the amount of other factors employed in the first position. The amount of other factors will depend on their costs; thus the supply curve of capital in terms of efficiency units is not independent of the supply curves of the other factors. In the simplest possible case, when the factors are uniform in nature, but falling in supply price, like the machines which become cheaper when more are employed, this difficulty does not arise, but in order to use this device for more complicated cases it is necessary to have a base line from which to start—some point at which the combination of the factors is known. For increases of output beyond this point the device will work accurately, but if the base line changes, all the separate supply curves of the factors have to be redrawn.

In some of the cases in which we have made use of this device there can actually be no base line. For instance, when we are comparing monopoly with competition, the proportions of the factors under monopoly (either producing a given output or working with a given number of men) may be different at every point from the proportions under competition. We introduced the separate supply curves of the factors in order to deal with the fact that the average cost curve of the commodity and the average net productivity curve of labour are not always the same under monopoly and competition. We now find that even the separate supply curves are not always the same under monopoly and competition. It was for this reason that, in the foregoing chapters, when we discussed economies of large-scale industry (shown by a falling supply curve of capital), we took as an example the case of machines becoming cheaper but unchanged in form, when more are employed by an industry, for in that case the supply curve of the factor is independent of the proportions in which it is used. It need not cause us much distress to discover that even the corrected comparisons between monopoly and competition are often inaccurate. There are so many general common-sense reasons why these comparisons should not be made (see Chapter 14) that we have not lost much when we discover this somewhat refined analytical reason why they cannot be made.

In the other cases where we have made use of this device it will not betray us. When we discuss the competitive demand curve for labour, we take as data the demand curve of the commodity and the supply curves (in natural units) of the other factors. We can then start at any point with the proportion of other factors (in natural units) to a given number of men, and then, taking this as the base line, construct the supply curves of the other factors in efficiency units for greater or smaller amounts. When we discuss the composition of the competitive supply curve, we must take as data the supply curves in natural units of all the factors. Then, starting from any output of the commodity, with the proportions of the factors that would be used in making it, we can draw up the supply curves of the separate factors in efficiency units for greater or smaller outputs.

Thus it is only in the comparisons between monopoly and competition, and then only in certain cases, that the above objection to our analytical device impairs its validity.

is a certain amount of money, this unit becomes more efficient when more is employed, because it can buy more machines as the machines become cheaper, and the supply price in efficiency units is falling. In more complicated types of economies of large-scale industry it cannot so easily be seen to which factor the economies can be attributed, but by means of this device they can be represented in the supply curve of any one of the factors, arbitrarily chosen.

When the amounts of the factors are measured in terms of efficiency units, constant physical returns will prevail. That is to say that when the amount of each factor in efficiency units is increased in the same proportion, output will also be increased in that proportion and the marginal physical productivity of each factor (measured in efficiency units) will be the same as before. Thus by means of this device conditions of constant physical returns are established, and any change can be imputed to the prices of efficiency units of the factors. This device for drawing up the supply curves of the factors throws no fresh light on the nature of increasing and diminishing returns, and can tell us nothing that we do not know already about the cost curve of a commodity. It is merely a piece of analytical apparatus which makes it possible to treat every type of increasing and diminishing returns in the terms appropriate to the simplest possible type, the type in which a uniform factor of production, composed of exactly similar men, acres, or machines, has a rising or falling supply price to an industry.

## 9

In the course of the argument in the foregoing chapters, we have made use of this device. When we drew up the demand curve for labour of a competitive industry we reckoned labour in natural units (men) and allowed economies to show themselves in a falling supply curve (in efficiency units) of the other factor (capital). If we wished to draw up a demand curve for capital, we should reverse the process and reckon capital in units of money and labour in units of efficiency, so that if there were economies of large-scale industry they would be shown in a falling supply curve of labour.

When we were dealing with the demand for labour of an individual firm, we found it unnecessary to make use of this device. We reckoned both labour and capital in physical units (men and money capital) and allowed the economies of large scale of the firm to show themselves merely in the increase of the physical productivity of labour and capital as the amount employed by the firm increases.

In comparing the demand for labour under monopoly and competition, we had to consider the relationships of the marginal productivity of a factor to the firm with its marginal productivity to the industry. One is the marginal physical productivity of the factor

to the firm multiplied by the price of the commodity; the other is marginal physical productivity to the industry multiplied by marginal revenue. It remains to show that we were justified in treating marginal physical productivity to the firm and to the industry as identical, so that the ratio of the marginal productivity of a factor to the firm to its marginal productivity to the industry is the same as the ratio of price to marginal revenue. If we were to reckon any factor not in efficiency units but in natural units, this would not be true. To take, once more, the simplest case in which capital consists of a certain type of machines which become cheaper (without any other alteration) when more are employed: then if we measure capital in money (which is the corrected natural unit), when one firm increases the amount of capital which it employs by one unit of money capital, machines become cheaper for all the firms, and if the amount of capital employed by the other firms measured in money remains constant, they are using more machines and producing a larger output. Thus marginal physical productivity to the industry would be greater than to the firm. But if we measure capital in efficiency units (in this case the machines, which are all alike) and if the only economy consists in the fall in the price of machines, then when one firm increases its employment of capital by one efficiency unit, a machine, and the other firms keep constant the amount of capital in efficiency units (that is, the number of machines), their output does not increase, and the whole benefit to them is shown in the fall in the price of machines. The marginal physical productivity of capital, measured in efficiency units, is then the same to the firm and the industry.

More complicated cases can be treated in the same way. If the number of efficiency units of capital employed by the other firms remains constant when the amount employed by one firm increases, then (by the definition of an efficiency unit) their output remains constant, and marginal physical productivity to the firm and to the industry are identical. The benefit to the industry due to the increase in capital is shown entirely in the cheapening of the efficiency unit of capital, that is to say, it is shown in the supply curve of capital to the industry, and not in the physical productivity of capital. Thus, when the whole of the economies are represented in the supply curve of the factor whose marginal productivity we are measuring, the marginal physical productivity of that factor is the same to the firm and to the industry. If the economies are shown in the supply curve of some other factor, this will not be the case. If economies are shown in the supply curve of capital, the marginal physical productivity of labour to the industry will be greater than to the firm.

10

We find that it is possible to represent both increasing returns and diminishing returns in the supply curves of the factors to an industry and from the point of view of a single industry they are perfectly symmetrical. Diminishing returns arise from a rise in the efficiency cost of a factor when more is employed, and increasing returns arise from a fall in the efficiency cost of a factor when more is employed.

But in their nature, as we have seen, increasing and diminishing returns are not symmetrical. Increasing returns arise when the employment of more of a factor has a favourable reaction upon the efficiency of the units already employed, and diminishing returns arise when the employment of more of a factor has an unfavourable reaction upon the price of the units already employed.

A type of increasing return symmetrical with diminishing returns would arise if a factor became cheaper (its efficiency remaining the same) when more was employed. This is very unlikely to occur in practice.[1] A type of diminishing returns symmetrical with increasing returns would arise if a factor became less efficient (its price remaining the same) when more was employed. This may sometimes occur. We found that increasing returns to an industry would arise in three ways. Firstly, it can arise from specialisation of firms. It is impossible to find a type of decreasing returns symmetrical with this. Secondly, it can arise from external economies which are independent of the size of any subsidiary industry, for instance from an improvement in the natural gifts of the labour force when a larger number of men are employed in one industry. If it were the case that when a large labour force was devoted to a single industry the labour deteriorated, so that each man became less competent when more were employed, we should have an external diseconomy symmetrical with this type of external economy. Thirdly, external economies can arise when a subsidiary industry becomes more efficient as it grows in size. External diseconomies symmetrical with this type of external economies are more likely to occur. If a machine-making industry were working under increasing cost, the supply price of machines would rise, and the same amount of capital, supplied at the same rate of interest, would buy fewer or worse machines. This would have the same effect from the point of view of the industry as if the supply price of capital rose when more was employed. But we must not

[1] A reduction in piece-rates may sometimes lead to an increase in the supply of labour, since each man may produce more pieces when he is paid less per piece. But this does not provide a true example of a falling supply curve of labour, since here it is the fall in the price of labour which is the cause of the increase in supply, and not the increase in supply which is the cause of a fall in price.

leave the matter until we have inquired why the machine-making industry is working under increasing cost,[1] and this must be due to a scarce factor of production somewhere, or else to the somewhat improbable cause of an actual deterioration of factors, supplied at the same price, as in the case where we imagined that workers became less competent when more were employed. Thus we find that the common types of increasing and diminishing returns are not symmetrical, but that it is possible to imagine cases in which the common type of diminishing return (due to a scarce factor) would be symmetrical with a rare type of increasing return (when the factor becomes cheaper as more is employed), and in which the common type of increasing return (due to improvements in the efficiency of the factor) is symmetrical with a rare type of diminishing return (when the factor deteriorates as more is employed). In any case from the point of view of an industry increasing and diminishing returns are perfectly symmetrical.[2]

Although from the point of view of an industry the various types of diminishing returns and of increasing returns can be regarded as symmetrical, the distinctions between them are of fundamental importance to society as a whole. A change in efficiency represents a net gain or loss to society as a whole, while a change in price does not. Thus changes in cost which are due to the rare type of diminishing returns and the common type of increasing returns (changes in the efficiency of the factors) are increasing or decreasing cost both from the point of view of the industry and from the point of view of society; while changes in cost due to the rare type of increasing returns and the common type of diminishing returns (changes in the price of the factors) are decreasing or increasing cost only from the point of view of the industry, and not from the point of view of society.[3]

[1] If the subsidiary industry is in a foreign country the chase may be conceived to end at the frontiers of the home country. Professor Pigou regards a rise in the price of imported raw materials (when the home industry expands) as an example of diseconomies of large scale to the home industry rather than as the result of the existence of a scarce factor of production. (*Economics of Welfare*, p. 222.)

[2] In the analysis set out in this book no account has been taken of decreasing cost due to a change in the price of the factors, or of increasing cost due to a change in the efficiency of the factors (measured in each case in terms of corrected natural units). But the analysis can easily be adapted to deal with these rare types of decreasing and increasing cost.

[3] See *Economics of Welfare*, pp. 219-27.

# INDEX

*Note.*—References to Professor Pigou's *Economics of Welfare* are to the Third Edition, and to Marshall's *Principles of Economics* are to the Seventh Edition, which is identical with the Eighth Edition.

THE END

*Printed in Great Britain by* R. & R. CLARK, LIMITED, *Edinburgh.*