On the Relations Between Cost and Quantity Produced†

BY PIERO SRAFFA

The Statical theory of equilibrium is only an introduction to economic studies; and it is barely even an introduction to the study of the progress and development of industries which show a tendency to increasing return.

Marshall, *Principles*, V, XII, 3.

I. *Description of the Problem*

IT may be said that there is today no economics textbook that does not contain a proposition of this sort: "We may, at a given moment, and with respect to a given market, divide all products into various classes. The first class will be made up of those commodities of which a larger quantity than that available at a given moment, in a given market, may be obtained *by a simple proportionate increase of cost*; the second class will comprise products which can be increased at *a less than proportional cost*. And finally the third class will consist of such products as cannot be increased at a given time and place except at *a more than proportionately increased cost*."[1] Whoever wishes to investigate which industries are to be found in one or other of these categories would find in the work of many writers the answer that 'agriculture' belongs to the third, 'manufacturing' to the second and industries that use more or less only direct labour would belong to the first category. For other, more recent writers, the solution is more complicated. Whilst, in general, leaving 'agriculture' in the third category, it is maintained that other industries can be found in any category, according to their particular circumstances. What these circumstances might be, from the point of view of the variation of cost in relation to variations in the quantity produced, has not been established, so that the curiosity of anyone wanting *[278]* to see the 'empty boxes' of constant, increasing and decreasing costs filled with concrete industries, remains more than ever unsatisfied.[2] But hope of reaching a

---

†*Translation of 'Sulle relazioni fra costo e quantita prodotta'*, Annali di economia *2 (1925): 277 - 328; by A. Roncaglia and J. Eatwell. This edition prepared by G. Langer and F. Lee, Chicago, Illinois. Page numbers of the original are shown in brackets.*

[1]Pantaleoni, *Principi di economia pura*, Firenze, 1889, pp. 225 - 226. [*Pure Economics*, English translation, pp. 187 - 188.]

[2]Clapham, 'Of Empty Economic Boxes', *Economic Journal*, 1922, pp. 305 ff.

classification remains even if it is put off until such time as 'there is available better statistical material' and until such time as men appear on the scene 'who have the qualities required for conducting a detailed intensive study of particular industries' and at the same time are 'versed either in the most intricate parts of economic analysis or in modern statistical technique'.[1] The hope should not be without foundation, in the light of the fact that, in the meantime, an important part of economic theory is based on the presumption that every industry ought to fit into one or another category, and every writer is careful to check if his conclusions apply to the three cases, and what the different consequences are in each case.

However, it remains to be seen if this presumption is well-founded; that is, if the absence of a classification of industries according to the criterion of the variability of cost is really due to the lack of data currently available and to the inability of scholars, or if, rather, the failing cannot be found in the very nature of the criterion according to which the classification should be conducted. In particular, it remains to be seen whether the *fundamentum divizionis* is formed by objective circumstances inherent in the various industries, or, instead is dependent on the point of view of the person acting as observer; or, to put it another way, whether the increasing and decreasing costs are nothing other than different aspects of one and the same thing that can occur at the same time, for the same industry, so that an industry can be classified arbitrarily in one or the other category according to the definition of 'industry' that is considered preferable for each particular problem, and according to whether long or short periods are considered.[2] These, in the first instance, are the problems we propose to discuss. *[279]*

The theory of decreasing productivity was always dealt with by classical writers in relation to the rent of land, and was therefore included, according to the traditional division of economics, in the theory of 'distribution'. Increasing returns on the other hand was discussed in relation to the division of labour, that is in the analysis of 'production'. But nobody, until comparatively recently, had thought of unifying these two tendencies in one single law of non-proportional productivity, and considering this as one of the

[1]Pigou, 'Empty Economic Boxes: A Reply', *ibid.*, p. 465.

[2]It may be helpful to emphasise once and for all that throughout this essay we are always dealing with long periods; which means to say, it is supposed that for every variation in the quantity of the commodity produced, a period of time is allowed that is sufficient to introduce all resulting modifications in the organisation of production, and the transitory effects that occur during the course of such adjustments before a new equilibrium is achieved are ignored.

bases of the theory of price. It could not have been otherwise, since greater division of labour was not generally conceived of as a phenomenon strictly dependent on the increase of the quantity to be produced, but rather was considered as an effect of progress in general. There was no evidence at all of that functional connection between quantity produced and cost of production, which is precisely what the law of non-proportional productivity consists of. It is true, however, that the law of diminishing productivity of the land gave prominence to a connection of this type, but recognition of the fact that greater output of necessity carried with it greater cost led only to consideration of the resulting variations in distribution. Moreover, this effect could not be considered a normal cause of variation of the relative price of individual commodities, for the increase in cost involved all, or almost all, commodities together, since almost all, in the final analysis, were derived from agricultural production[1] and hence the action of decreasing productivity increased proportionately the cost of each.

The idea of interdependence between quantity produced and the cost of production of a commodity produced under competitive conditions is not suggested by experience at all and could not arise spontaneously. It can be said that all classical writers accept implicitly, as an obvious fact, that cost is independent of quantity, and they do not bother to discuss the contrary hypothesis. This idea of interdependence has taken shape recently, in an indirect way, as the result of the change in the basis of the theory of value, from cost of production to utility. It should not be surprising that, while for a long *[280]* time people have continued to talk of cost as being independent of quantity produced, as soon as utility was subjected to a methodical analysis it was seen that of necessity the utility of a commodity depends on the available quantity of that commodity.

The 'demand function' is based on an elementary and natural hypothesis, that of decreasing utility. Whilst in production the functional relationship is the result of a much more complicated set of hypotheses. The fact remains that only *after* the studies of marginal utility had called attention to the relationship between price and quantity (consumed), did there emerge by analogy the symmetrical conception of a connection between cost and quantity produced.

The importance of the laws of variation of cost in relation to the determination of the price of single commodities has appeared only in consequence of the 'fundamental symmetry of the general relations in which

[1] *Cf.* below p. 41, note 1, on the meaning of the word 'corn' in the classics.

demand and supply stand to value'.[1] According to this doctrine 'the normal value of everything . . . rests like the keystone of an arch, balanced in equilibrium between the contending pressures on its two opposing sides. The forces of demand press on the one side, those of supply on the other; . . . .'[2] Such symmetry depends on the non-proportionality of the total cost of production to the quantity produced. If the cost of production of every unit of the commodity under consideration did not vary with variations in the quantity produced the symmetry would be broken; the price would be determined exclusively by the expenses of production and demand would be unable to have any influence on it at all.

It is on the basis of this position, that is from the point of view of the determination of particular equilibria of individual products under a regime of free competition, that we will examine the theoretical foundations of the laws of variation of cost.[3] *[281]*

II. *Increasing Costs*

The law of diminishing returns is defined in Palgrave's *Dictionary of Political Economy* with these words: 'If one, or more, of the industrial agents, the co-operation of which is necessary for the production of any commodity, be increased, the others remaining unaltered, the amount of the product will generally be increased. If the increase of the product be in a less proportion than the increase of the industrial agents considered, we express the fact by saying that in this case the product obeys the law of diminishing returns'.[4]

This definition is generally accepted and we can take it as the basis of our discussion of diminishing returns. However, before going further it is necessary to clear up a point that can cause confusion. The definition does contain the substance of the hypotheses that are characteristic of diminishing

---

[1]Marshall, *Principles of Economics*, 8th Edition, 1920, p. 820.

[2]*Op. cit.*, Preface to the 2nd edition, 1891.

[3]The variations of cost can be considered in relation to the quantity produced: (1) by a monopoly; (2) by a single firm in competitive conditions; (3) by the totality of competing firms. By occupying ourselves with this latter case, we will have occasion to examine also its relationship with the second.

[4][*Palgrave's Dictionary of Political Economy*] Vol. II, p. 583, under the heading *Laws of Political Economy*.

returns, which it is necessary to distinguish from those of a complete different nature which relate to increasing returns. But the manner in which it is expressed obscures such a distinction, to the point of making many believe that now one and now the other of the two modes of variation of productivity may be derived from the same conditions. In the *Dictionary* Palgrave falls into this confusion, when defining the 'law of increasing returns' he says: 'when *under the circumstances supposed above* (see 'Law of Diminishing Returns'), the increase of product is in greater proportion than the increase of the industrial agents concerned the *Law of Increasing Returns* is said to be in operation'.[1] It is necessary to point out that the 'supposed circumstances', which give rise to the variation of cost, according to the *Dictionary*, are the same in the two cases. The circumstances are that, if we consider, for simplicity's sake, only two factors, one remains constant while the other increases. This presupposes: (a) a modification in the *proportion* between the quantities of the two factors; (b) an increase in the *size* of the industry. Now it is obvious that the connection between the two circumstances is purely fortuitous, and depends on the fact that variation of the proportion *[282]* between the factors derives from keeping one of them constant while the other increases. It is exclusively the first circumstance (a) that gives rise to decreasing productivity, *notwithstanding* the influence of the second, which can operate in the opposite direction. Increasing productivity stems only from the second circumstance (the increased size of the industry, which could obviously also derive from the increase of *all* factors of production) *notwithstanding* the first.

The identity of the conditions that give rise to the two opposing tendencies is therefore illusory. This illusion derives from too literal an interpretation of the expression 'constant factor', by considering such a factor to be susceptible neither to increase nor to decrease. But in general it is arbitrary to suppose that if there is an excess of one factor it is not possible to get rid of it. In reality it is usually found that the 'constant' factor cannot be increased, but that it can indeed be reduced.[2] The typical case of a constant factor is land. The theory of rent is based on the fact that the amount of land is given. But consideration of the spread of cultivation from the best land to the worst land,

---

[1] *Loc. cit.*

[2] Diminishable, of course, at the wish of the person using it. But as to the effects on the theory of rent, we cannot agree that the 'constant' factor is diminishable at the wish of the person who is providing it, for that would result in the possibility of using it in a different way, and thus the rent would, from the point of view of the industry considered, be turned into cost.

shows that no-one thinks that the farmers are always compelled to cultivate the whole of the existing surface. However, it is on precisely this supposition that the identity claimed between the conditions that we are now examining has been based. This can already be found in the formulation of the law of decreasing productivity first given by Turgot: "Seed sown on land that is naturally fertile, but totally unprepared, would be an advance almost entirely wasted. If the soil were tilled once, the produce would be greater; tilling it a second, a third time, might not merely increase the produce two or three times, but four times or ten times. The produce would thus be augmented in a much larger proportion than the advances increased, and this up to a certain point, at which the produce will be as great as possible compared with the advances. Past this point, if the advances are still increased, the produce will still increase, but less and less, *[283]* until the fertility of the soil being exhausted, and art unable to add anything further, an addition to the advance will add nothing whatever to the produce."[1]

This passage is noteworthy, not only for the originality of its contents, but also for the precision of expression. However, in the first part, in which he maintains that there is a tendency to increasing productivity from the first 'doses of capital and labour' applied to a given price of land, he is stating only what would happen in the case of a farmer who had limited resources and did not know the best way to use them. It is obvious that if the farmer knew the best way, instead of sowing and tilling *all* the land once, it would be better for him to sow and till, say, *half* the land three times, because in this way he would obtain a quintupled product. More precisely, he ought to cultivate a quantity of land such that his resources would allow him to carry cultivation to the point of maximum productivity. If the problem that he has to solve, rather than being that of obtaining the maximum product with a given quantity of capital and labour, was to obtain a given product at a minimum cost, the solution would be analogous: he would have to make use of only that amount of land, which, cultivated up to the point of maximum productivity, gave him the product required. This holds good, of course, until he has to put all the land available to him, which we are assuming to be uniform quality, under cultivation. Up to this point productivity would be constant, that is, the product proportional to the expense, since with the growth of the expense, the quantity of land cultivated would grow in equal proportion. This can clearly be shown with a diagram (see Fig. 1). We represent on the axis Ox the successive doses of 'capital and labour' which are used on the whole of a given piece of land, and we indicate with their ordinates the product obtained by

[1] 'Observations sur le Memoire de M. de Saint-Peravy en faveur de l'impot indirect' in *Oeuvres de Turgot*, Paris, 1844, Vol. I, p. 421.

each dose. The curve OAB, thus defined, which we call *the marginal productivity curve*, represents a state of affairs similar to that described by Turgot. If instead of measuring the increment of product due to the addition of a dose of capital, we were to represent on the ordinate the total product of the doses divided by their number, we obtain the curve OPD, which we call *the average productivity curve*. The two curves are related in such a way *[284]* that, if for any given point Q on OAB, we take, from point R of equal abscissa, the normals to Ox and to Oy, the rectangle ORTS is of equal area to OQS. Point P, the intersection of the two curves, corresponds to the maximum ordinate of the curve OPD[1] and is the point of maximum productivity indicated by Turgot.

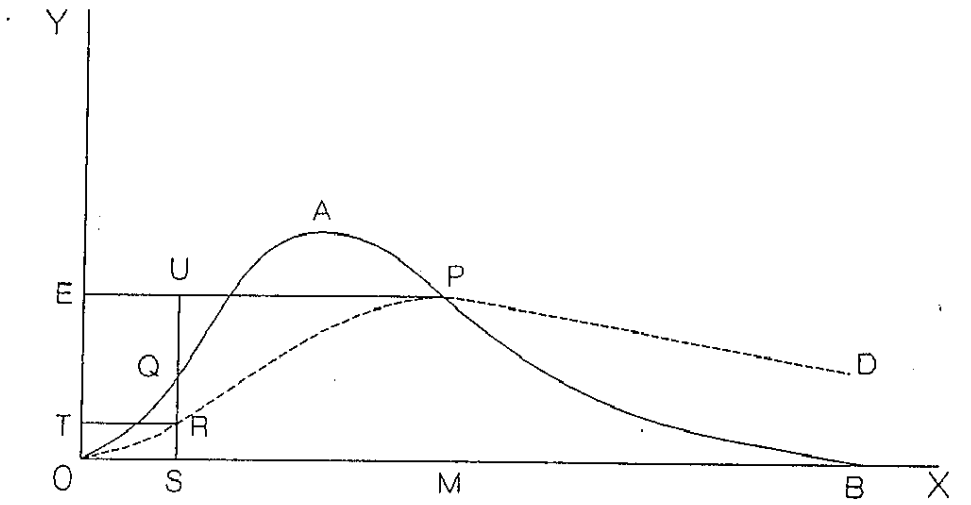

Figure 1.

These curves represent the productivity conditions on a given quantity of land. If we suppose that the land is homogeneous over all its surface, we can obtain for every fraction of the surface, two curves analogous to those in Fig.

[1]This property becomes obvious when the curve is considered as discontinuous; that is, it is supposed that the quantity of capital and labour increases through finite increments. In such a case the average product of any quantity of capital can be obtained from the weighted mean of the average product of the quantity which is immediately inferior and the (marginal) product of the increment. Therefore, since in Fig. 1 a *maximum* of the mean product (PM) corresponds to the quantity OM, the marginal product of the quantity immediately inferior to OM must be greater than PM and the marginal cost of the quantity immediately above must be less than PM. Therefore the two curves will intersect each other at the point P. For an analytical demonstration of an analogous case, see Edgeworth, 'Contributions to the Theory of Railway Rates, IV', *Economic Journal*, 1913, p. 214.

1. The points of each of such pairs of curves, in relation to the corresponding points of the curves for the whole of the land, will, for abscissas proportional to the fraction of land to which each pair refers, have equal ordinates. Therefore, for each piece of land, maximum productivity will be equal to MP. *[285]*

Returning to Fig. 1, it follows from this that none of the points of the two curves with abscissas less than OM can be a point of equilibrium. If the cultivator decides to use a quantity of capital and labour (for example OS) less than that necessary to bring the cultivation of all the land to a state of maximum productivity, it is better for him not to cultivate all the land obtaining from it a product OTRS, but to cultivate that part of the land which, with the same capital and labour, attains maximum productivity and gives him the product OEUS. (That is, precisely the part of the land that stands to the whole of the land in the same proportion in which OS stands to OM). Therefore, with the increase of the capital and labour used in cultivation, the productivity curve will be represented by a straight line EP up to the point of maximum productivity, and only beyond this point will the curve begin to descend. Over the whole length of the curve productivity may be constant or decreasing, but in no case can it be increasing.

What has been said has presupposed the condition that the factor whose maximum quantity is assumed to be 'constant' is indefinitely divisible and therefore that it should be possible to use only a part of it in production. In general, there is no reason to suppose that the quantity of this factor in existence is also the least that is always necessary for production, and therefore there will always be at least a certain range within which the quantity may be conveniently varied. But it is possible that, below a certain limit, the quantity of the 'constant' factor cannot be reduced without leading to a diminution of product even greater than that caused by applying a smaller quantity of other factors to an identical quantity of the 'constant' factor. However, this occurs only when production is very small;[1] and the *[286]*

[1]Strictly speaking we can in each case ensure that the other factors applied to the constant factor give an increasing productivity, by reducing the quantity of the 'constant' factor to the extreme limit. From this point of view, therefore, Edgeworth's argument: 'there is, in one respect, a greater unity in the action of Diminishing Returns—that it always rules, *provided that we take sufficiently large doses*', is not justified. ('Railway Rates, II', *Economic Journal*, 1911, p. 522). Such unity could also be found in the action of increasing productivity, since it always works *on condition that sufficiently small quantities are considered.* The same objection can apply to Pigou, who, after having announced this 'law of diminishing returns to individual factors of production' states that 'there is no law of increasing returns to individual factors corresponding to it'. (*Economics of Welfare*, London, 1920, p. 704.)

smallest area of land that is adequate for cultivation is certainly small enough to be regarded as irrelevant with respect to a large volume of production, for example that of an entire State. And this is precisely the respect in which one ought to consider any one factor—in our example, land—as constant. In fact, from the point of view of a single producer (whose production would be so small as to enable him to achieve this type of increasing productivity if there were for him a constant factor) land is no more difficult to increase than other factors, and with the same means by which he obtains an increase of those factors he can always obtain a larger amount of land.[1] But in reality, even in the case of the single producer, increasing productivity, which always occurs, at least initially, usually has entirely different origins from that considered here. With a view to making the distinction—to which we will revert (below p. 23)—clearer, it is necessary to mention that initial increasing productivity from successive doses of a factor A applied to a constant quantity of another factor B, presupposes the *negative* productivity of factor B. This means that if, inverting the terms of the problem, we were to consider the productivity of successive doses[2] of factor B applied to a constant quantity of A a point would be reached at which the addition of further doses of B would cause a *diminution* of total product (and not only of the marginal product, which would become negative). Let Fig. 1 represent the condition of application of capital and labour to a constant quantity of land k. Let us suppose that k is so small as to be indivisible without loss of efficiency in cultivation, and therefore the average productivity curve cannot *[287]* conform to EP in the initial section, but must rise as OP. Further, let us call c the quantity of capital and labour that, on the same amount of land k, gives the maximum average product per unit of capital and labour (that is, OM). If we now imagine another diagram, which, because of its simplicity it would be pointless to draw, in which the abscissas represent successive uniform doses of the same land used together with a constant quantity of capital and labour (which we will take to be equal to c), and the ordinates are the product

[1] *Cf.* below p. 22.

[2] The expression 'productivity of a factor' can be misleading. It is therefore useful to clarify that by average product of a factor we mean the total quantity of the product divided by the number of units of that factor which, together with others, it is necessary to use in the production of that quantity; and by marginal product of a factor we mean the increment of product that is obtained by adding to a given quantity of factors a 'dose' of the factor being considered. It is a question of an analytical expedient, which does not in the least imply that the factor under consideration contributes more or less to the product than the factors with which it is combined. Given these definitions, the propositions that follow are not exposed to the criticisms that Loria directs at this expression (*I fondamenti scientifici della riferma economica*, Turin, 1922, Chapter 1).

obtained with the addition of each dose. This curve will be descending for all its length. When the quantity of land used in totality reaches the quantity k, the curve will cut the axis of the abscissas and its ordinates will become negative for every subsequent dose of land used (which is to say that every subsequent dose will *destroy* a part of the product).[1]

It is clear that, since it is being assumed that the factors are being used in the best way, once that point is reached the land would cease to be increased, even if it were free, because the best way of using a further dose of the land would be, precisely, not to use it. Therefore, the type of increasing productivity that we are considering, deriving from the fact that the proportion between the factors is at the outset unfavourable, happens only when a factor exists in an excessive and harmful quantity, and it is not possible to get rid of it without cost.

Having specified the hypothetical conditions under which occurs the phenomenon of decreasing productivity, considered as a general fact connected with the proportion in which different factors of production are combined, it is appropriate to investigate whether there is a common cause that produces such a uniform effect in very different fields of production. It is surprising that most writers are agreed in searching for the particular circumstances of the various cases in which diminishing returns occur. Some go so far as to object that 'the causes are too diverse to allow us to talk of a law of decreasing returns'. Edgeworth, whilst opposing this extreme opinion, holds that 'with respect to Diminishing Returns in the sense which is of particular interest in the railway industry, I think we may say that the phenomenon has all manner of causes except those botanical ones which are characteristic of the law in its first *[288]* and still most important form relating to agriculture'.[2] And Marshall states that 'the tendencies of diminishing utility and of diminishing returns have their roots, the one in qualities of human nature, the other in the technical conditions of industry'.[3] Faced with these explanations of the tendency towards decreasing productivity, which claim to find the reasons for it in the peculiar circumstances of every single case, the question springs to mind, is it not very strange that two such heterogeneous

---

[1] *Cf.* Carver's 'first case' in 'Diminishing Returns and Value', *Scientia*, II, p. 338.

[2] [Edgeworth] 'Railway Rates, II', *Economic Journal*, 1911, pp. 552 - 553. *Cf.* the quotations there that '*well illustrate the variety of causes leading to a similar result in different departments of production*'. [Sraffa's italics]

[3] [A. Marshall] *Principles*, 8th edition, p. 170, note 20.

things as human nature and industrial technology should bring about results so similar? And it is not just a question of two single elements. It is even more improbable that these 'technical conditions' which cause diminishing returns of the subsequent doses of a factor applied to another which is constant, should be similar in a large number of very different industries, and even in the 'production' of utility through the consumption of the commodities. If these industries resemble one another in diminishing returns to a factor, it is more likely, and simpler, to assume that this resemblance is due to the single element that they have in common, that is, their relationship with 'human nature', and that this should be sufficient to impress on them this common characteristic.

This explanation presupposes two conditions: (1) the application of the principle of substitution, that is to say the criterion by which economic choice is made; (2) the existence of a certain degree of variety and of independence among those elements that make up the variable factor, or between those parts that make up the constant factor, or between the methods by which two factors can be combined, (that is, between the ways in which the variable factor can be used). Given these conditions, diminishing returns must of necessity occur because it will be the producer himself who, for his own benefit, will arrange the doses of the factors and the methods of use in a descending order, going from the most favourable ones to the most ineffective, and he will start production with the best combinations, resorting little by little, as these are exhausted, to the worst ones. The complicated nature of the contrary hypothesis based on 'technical conditions', *[289]* is a major argument against it, for it implies the supposition that for each industry there exists an independent law of diminishing returns. Moreover, it is very difficult to check to what extent diminishing returns are based on particular cases, for it is difficult to find an industry in which no possibility is left for substitution. However, if in given circumstances, there was a material necessity to resort to successive productive combinations in an order predetermined by non-economic considerations, there would generally be no reason why they would follow a decreasing order of efficiency rather than an increasing order.

We will take the case of agriculture, since, just as the generalised law of diminishing returns had its origin in agriculture, so the general explanation of that law based on the 'technical conditions' has developed from the explanation based on agricultural technology. J. S. Mill was the first to point out that 'the decreasing ratio in which the product of the soil is increased by an increased application of labour' is one of 'those truths which political economy seems to borrow . . . from the physical sciences to which they

properly belong'.[1] Such an assertion has been accepted without argument by many writers,[2] and even Pantaleoni wrote that 'this so-called law . . . in reality is simply a premise of economic laws',[3] and, more precisely, it 'is a datum of agrarian technology';[4] 'the demonstration of this so-called law must either be obtained from the examination of facts, or be replaced by the transformation of the law into a postulate or hypothesis'.[5] This implies that agricultural technology enforces the manner in which each of the successive increments of expenses of production must be used on a given piece of land, and through a set of fortuitous circumstances that are unknown to economics, determines that the product of every equal and successive outlay should be a decreasing one. But the facts are otherwise. When, having spent an annual sum on the cultivation of a given land, and wishing to spend another thousand lire, reference to the agricultural technology will indicate *[290]* not only one way but a whole series of different ways, A, B, C, D, . . . , in which it is technically possible to spend the additional 1,000 lire. It will be possible to buy additional fertilizer, or make a deeper ploughing, or improve the quality of the seed, or one hundred other possible expenditures, or any combination of these. In addition, the technology will determine that by spending the 1,000 lire on method A a product x_a will be obtained, by spending the 1,000 lire on method B, a product x_b, etc. Beyond this point the farmer will no longer be guided by technology, and he will select, on the economic criterion the method which will give him the largest product from the methods of using the 1,000 lire. This choice is already, in itself, a long way from agricultural technology, and it will be even further from it if x_a, x_b, ... are quantities of heterogeneous products that to be compared must be reduced to the common standard of their value. Let us suppose that the choice is made to spend the 1,000 lire on method B. If, subsequently it is decided to spend another 1,000 lire the choice will be restricted. There will no longer be either method B, or those methods among the others that are incompatible with B, that is that can no longer be used when B is used. This will leave the choice, let us say, between methods A, C, D, . . . , each of which *in the preceding conditions*, (when the 1,000 lire had not

---

[1][J. S. Mill] 'On the Definition of P. E.' (1829) in *Essays on Some Unsettled Questions*', p. 133, note.

[2]See, for example, Cairnes, *Logical Method of P. E.*, p. 34, J. N. Keynes, *Scope and Method of P. E.*, p. 85, etc.

[3][Pantaleoni] *Principi di economia pura*, p. 224 [English translation p. 186].

[4]*Ibid.*, p. 10. [English translation p. 4, but here the translation is ours.]

[5]*Ibid.*, p. 224. [English translation p. 186].

yet been spent on B), would have given a product less than, or, at best, equal to that of B. If, in the current conditions, after having spent 1,000 lire on B, the productivity of these methods is unchanged, (which is the case when they are perfectly independent of the use of method B), it is clear that the second 1,000 lire will give a product less than the first 1,000 lire, since the producer has chosen and has acted in precisely such a way as to make this happen. If the return from the remaining uses, in the new conditions, were diminished, we would have a case of a 'physical law of diminishing returns' and the result would take place *a fortiori* through the economic law coinciding with the physical law. Finally there remains for consideration the case in which, after having used method B, productivity of the other uses is increased. Now this, which should be a case of increasing productivity, cannot happen unless the cultivator has made a mistake in his calculations. If this case occurred, instead of spending the preceding 1,000 lire on method B, he should have spent it on a mixed method M, (which agricultural technology would certainly have indicated), comprising, let us say y lire used on method B and *[291]* 1,000 - y lire on method D, applying method M to half his land. Then he would have been left with the possibility of using another 1,000 lire on method N, identical to M, to be applied to the other half of his land. This case comes back to that considered above, page 6, for which, when a second ploughing increases the product more than the first, it is better to plough half of the land twice, rather than plough the whole of the land once. Here, too, we can have increasing returns only in the case in which the land being considered is so small as for it to be impossible to subdivide it for cultivation, without loss of product. But, leaving aside this extreme case, which is generally irrelevant, it may be maintained that given the assumption, the increase of only a few of the factors of production usually adds to the product in a decreasing or, at best and for a short time, in a constant proportion.

When the law of diminishing returns is considered from this point of view, it can clearly be seen why Ricardo preferred to emphasise the loss of productivity arising from the gradual extension of cultivation to land less and less fertile, leaving in the background the loss of productivity deriving from the application of successive doses of capital and labour to one and the same piece of land. The proposition that the productivity of a given piece of land is to a large extent *independent* of whether or not another piece of land is cultivated is both true and obvious. But the productivity of a given dose of capital applied to a piece of land is to a much lesser extent independent of whether or not another dose of capital is applied at the same time to the same piece of land. Thus the truth and generality of the law of diminishing returns is much greater if it is based on the variety of the pieces of land, than

if it is based on the variety of the doses of capital and labour, or on the variety of purposes for which equal doses can be used.[1]

The characteristic of Ricardian theory which we have *[292]*[2] identified as fundamental, that is, attributing diminishing returns to an economic rather than a physical cause, has been very ably criticised by Wicksteed. He begins by dividing productivity curves into two categories: *descriptive* curves and *functional* curves. This distinction broadly coincides with the contrast between an *economic* law and a *physical* law of diminishing returns which we discussed above. Wicksteed constructs the descriptive curve, which represents the most important Ricardian type of diminishing returns, in the following way: 'different qualities of land are represented along the axis of X, and their supposed relative fertilities to a fixed application of labour and capital along the axis of Y. The 'marginal' land will occupy the extreme place to the right. This is not a functional curve; for the height of Y does not depend upon the length of X, the unit being expressly so placed on OX as to produce a declining Y. It is applicable to land or to anything else of which typical units can be arranged in ascending or descending order of efficiency'.[3] The functional curve is defined as follows: 'take a given fixed area of land of a certain quality and consider what would be its yield if it were 'dosed' with a certain quantity of labour and capital represented by a unit on the axis of X. Increase the dose till a further increment of labour and capital would not produce as large an increment in the yield of this land as it would if applied to some other piece of land of the same or different quality, or if turned to some non-agricultural business. The last increment actually applied is the 'marginal' increment, and it measures the distributive share of a unit 'dose' in the product'.[4]

---

[1]This was exactly the reason why Ricardo, having analysed as distinct the two forms of the law (see *Works*, McCulloch's edition, especially the note on p. 251) prefers to make use predominantly of the first for successive deductions; as is confirmed by the fact that, while he represents the passage from better land to worse land as a true and obvious fact; he speaks of decreasing productivity on a given piece of land as a probable, not a certain thing, prefacing the exposition of the second form with conditions such as 'It often, and indeed, commonly happens....It may perhaps be found....' (*Works*, p. 36).

[2]*[Editor's note: This page is incorrectly numbered 291 in the original.]*

[3][P. Wicksteed] 'Political Economy in the Light of Marginal Theory, etc.' *Economic Journal*, 1914, p. 17.

[4]*Ibid.*, p. 18.

Thus the basis of the distinction is this: in the descriptive form the position in the sequence held by each dose is determined by the productivity of the dose; this productivity is therefore independent of the number of doses utilised. In the functional form, however, it is the place held by each dose which determines the productivity of that dose; this productivity is therefore strictly dependent on the number of doses utilised. In *[293]* other words, in the first form it is assumed that all the doses considered are distinct from one another; and thus even if utilised under the same circumstances, they have different productivities. In the second form, the nature of all doses is assumed to be the same, but the doses have different productivities due to the differing circumstances of use. In both conceptions we speak of a *marginal* dose, but, as Wicksteed points out, the expression has 'entirely different senses'. In the first case, it is a particular dose, that of lowest quality; in the second it may be any one of the doses. In the latter case, 'it is not any peculiarity of the 'marginal' increment that makes it yield less than the others. It does not. They all have exactly the same differential effect on the yield as to which none is after or before the other. The height of this differential or marginal yield is dependent not upon the nature of each several dose, but upon their aggregate number.[1] Now, of these two types of curves and *margins*, Wicksteed rejects the first 'which neither illustrates nor proves anything except that the better article commands the better price',[2] since it results from an arbitrary ordering, and in consequence he denies that the Ricardian theory of rent based on it has any value. As for the second case, he accepts it as the foundation of the 'differential theory of distribution', on condition that it is applied not only to land—the remuneration of which would be determined in the same manner as that of other factors of production—but to all factors. We cannot dwell here on the use which Wicksteed makes of his distinction in relation to the theory of distribution. Nor can we linger on the objections he raises against the determination of market price by the intersection of *[294]*

[1] *Loc. cit.*, p. 18.

[2] *Common Sense of Political Economy*, 1910, p. 572. 'And in very truth that is all the Ricardian law of rent amounts to.' (p. 569). Wicksteed seems to hold that the relative superiority of the units of the factors are a datum of the problem. This would be true if the factors were all homogeneous, and the theory of distribution would be reduced to stating that the return to every factor is exactly proportional to its size, since this is precisely what the superiority would come to. But this is not so, and to determine this superiority is precisely one of the *quaesita* of the theory. Relative superiority cannot be fixed as an absolute criterion, but varies with the conditions of production. Thus, for example, as Marshall has shown (see below, p. 17), of two pieces of land A and B, which, when cultivation is light, A obtains the larger rent, it can happen that , intensively cultivated, B gives the greater rent. Which is then 'the better article' of the two in absolute terms?

supply and demand curves, to the extent of maintaining that the supply curve does not exist[1] and that it is necessary to consider, as determinants of price, only the quantity of the commodity in existence and the demand curve ('It is a curve representing a function').[2] We must restrict our considerations to the distinction itself. We will emphasise that, on the basis of what has been said in the preceding pages, the distinction appears to be groundless. Any decreasing curve with a general and not merely an accidental character, must be a 'descriptive curve'. Note that in the case of the functional curve, according to Wicksteed, 'the height of this differential or marginal yield is dependent not upon the nature of each general dose, but upon their aggregate number'. But this proposition is incomplete, for if it is true that the doses are identical and yet give a different yield, this implies that they are put to different uses, and therefore the product of the marginal dose is dependent precisely on the nature of its use. Therefore, in the functional curve the productivity of the marginal dose does not depend directly on the aggregate number of doses, but rather because the previous doses having already been applied to the best uses, a less productive use is left for the last dose. And the larger the number of doses the lower must we descend along the descending hierarchy of available uses. This hierarchy belongs to the genus of descriptive curves, for the uses have been placed 'arbitrarily' in decreasing order, and not because of any material necessity. Thus the 'functional curve' merely transfers the 'difference in nature', and therefore the 'arbitrary' ordering from the doses themselves to their modes of utilisation. But the relationship that links the number of doses to marginal productivity, that is, the productivity curve, belongs, in both cases, to one genus alone.[3] Obviously *[295]* (and this can be said in both cases) the arbitrariness is not, as

---

[1] *Op. cit.*, *Economic Journal*, 1914, p. 13.

[2] *Loc. cit.*, p. 12.

[3] We have criticised the distinction introduced by Wicksteed only from the point of view with which we are concerned, the nature of diminishing returns. It would be objected that the distinction between descriptive curves and functional curves is fundamental as far as the theory of distribution is concerned. In fact, in the first construction, the (different) pieces of land successively placed under cultivation in decreasing order of fertility receive different remunerations, while in the second the (equal) doses of capital successively employed on a given piece of land all receive, at each moment, the same remuneration. It would therefore seem that, according to whether the diversity (from which diminishing returns is derived) is found in the doses themselves or in the way in which equal doses are used, as *gratuitous factors* of production (according to Edgeworth's conception, 'Railway Rates, I' in *Economic Journal*, 1911, p. 357), to recognise that if these gratuitous factors were appropriated (for example) the return that they would receive would be different and, just as with different pieces of land, would be proportional to their efficiency.

Wicksteed seems to think, on the observer's part—who would arrange his pieces of land in decreasing order just as he would arrange a row of men in order of size.[1] But it is on the part of the producer himself, who, in effect, only uses his freedom to behave in the manner most rewarding to himself.

The same argument may be repeated for the case of diminishing utility (and therefore for the demand curves derived from it) which is a special case of diminishing productivity, when we consider utility as product, the commodities consumed as the variable factor of production, and the 'sensitive organism' as the constant factor.[2] It is not any allegedly psycho-physical law which endows diminishing utility with generality, but the possibility of using different doses of a commodity to satisfy different needs and the desire to utilise the first doses to satisfy the most urgent needs.

Having examined the objection that the decreasing order of fertility in which the various pieces of land are arranged is arbitrary, let us go on to consider another objection—the denial of the possibility of classifying the pieces of land according to their fertility, such that the ordering does not change with the increase in the intensity of cultivation.[3] It is clear that if this were true, the construction of the static curve of diminishing returns, based on the order of fertility of the pieces of land, would no longer be conceivable. Such an objection is important, not only from the point of view of the application of the theory to agriculture, but also from the point of view of the 'universal law of diminishing returns', which concerns us here. If the objection were well-founded, it could easily be extended to the criterion for judging which is the best use as between different uses of a given increment *[296]* of a factor, or which is the best use between various doses of any factor, each having different qualities. Marshall says that Ricardo expressed himself 'carelessly as though there were an absolute standard of fertility', when he stated that with the growth of population, pieces of land of even poorer quality are gradually put under cultivation. Marshall dedicated a paragraph of his *Principles*[4] to the demonstration of this point. 'There is no absolute measure of the richness

---

[1][P. Wicksteed] *Common Sense etc.*, p. 539.

[2]*Op. cit.*, p. 570.

[3]This objection does not, of course, refer to those variations which are outside the scope of the discussion: variations in the relative fertility of several pieces of land, that derive from modifications in technical knowledge, in systems of cultivation, and in the nature of the harvest.

[4][A. Marshall, *Principles*] Book IV, Chapter III, paragraph 3.

or fertility of land. Even if there be no change in the arts of production, a mere increase in the demand for produce may invert the order in which two adjacent pieces of land rank as regards fertility. The one which gives the smaller produce, when both are uncultivated, or when the cultivation of both is equally slight may rise above the other and justly rank as the most fertile when both are cultivated with equal thoroughness'.[1]

The question to be resolved is this: what is the definition of fertility (in the generic sense of 'superiority') that ought to be adopted when arranging the pieces of land in the order in which it is best to place them under cultivation? The possible definitions, and hence the ones accepted by different writers, are very diverse. Marshall considers the more fertile of two pieces of land, in a given equilibrium situation, (that is, such that the product of the marginal dose of capital used is equal on both pieces of land),[2] the one which gives the larger average product. This criterion leads to the conclusion that, with the growth of the intensity of cultivation, the order of fertility of the pieces of land changes. The same thing occurs with other definitions–for example that of Malthus: 'land of an inferior quality requires a greater quantity of capital to make it yield a given produce';[3] or the definition of J. S. Mill: 'inferior land is land which with equal labour *[297]* returns a small amount of product'.[4] These two definitions have, moreover, the inconvenience of presupposing that two pieces of land that are being compared are of equal area. But for this, one would have to consider the more 'fertile' of two identical pieces of land, the one with largest area. Now, the attribute of extent is certainly the basic attribute of land, but it has nothing to do with the definition of fertility that is required for the first type of diminishing returns; for there is no necessity to suppose that pieces of land successively put under cultivation are of equal

---

[1] *Op. cit.*, p. 157.

[2] p. 160. It could be thought that Marshall implies instead that the pieces of land are cultivated with the same amount of capital; but this condition will generally be incompatible with the other one, that the marginal productivities should be equal on both the pieces of land. This land condition seems better to reflect 'equal intensity of cultivation' and to be more consonant with the context of the passage quoted on p. 157, and, in general, with paragraph 3, which implies the existence of successive states of equilibrium on the market. *Cf.* especially the diagrams on p. 158, note.

[3] [T. R. Malthus] *An Inquiry into the Nature and Progress of Rent,* 1815, p. 27.

[4] [J. S. Mill] *Principles of P. E.*, Book 1, Chapter XII, paragraph 2; in VII edition, Vol. 1, p. 221.

area.[1] In the light of this fact, these definitions would lead us to the absurd conclusion that, *ceteris paribus*, pieces of land of the largest area are cultivated first. These, and other definitions that could be given, have the advantage of being fairly close to the vague conception that is commonly held of 'fertility'. But what we need is a criterion that indicates the order in which it is best to cultivate successively different pieces of land, and which *in every case* would hold good, independently of the subsequent wish to take cultivation to a greater or lesser degree of intensity. Now, it is best to cultivate first of all—and must therefore be considered the most 'fertile'—that piece of land which, at the point at which its marginal productivity is equal to the average productivity, has a productivity greater than all the other pieces of land. Referring to Fig. 1, it is that piece of land the curve of which has at the point P the highest ordinate PM.[2] That this would be the criterion followed in practice derives from the fact that in the cultivation of every piece of land, at least the point of maximum average productivity should be reached, and only after this would one pass on to another, less fertile, piece of land.[3] Therefore, if one were first to cultivate another piece of land, one would receive a smaller product for every unit of expense. The order of fertility thus determined does not change with the intensification of cultivation since the *[298]* form of the two productivity curves, and therefore the position of their point of intersection, does not change with a change in the index M.

We believe this analysis has shed sufficient light on the essential character of diminishing returns, in that diminishing returns derives from it being desirable and generally possible to arrange the efficiency of the doses of the factors of production and the different ways of using them in descending order—an ordering that is determined exactly. We now examine a case in which the principle has been wrongly applied. Barone wanted to extend it to the supply curve of a product under a regime of free competition. Having seen that 'there co-exist on the market entrepreneurs producing the same product at different production costs'[4] he classifies them in increasing order

---

[1] *Cf.* Edgeworth, 'Railway Rates, I', *Economic Journal*, 1911, p. 353.

[2] This definition includes the extreme case in which productivity is decreasing right from the start, since the two curves would have in common only the initial point, and the first piece of land to be cultivated would be the one with the greater initial productivity.

[3] Of course the first form of decreasing productivity ignores, but does not quite exclude, the possibility that before passing on to a second piece of land, the cultivation of the first is intensified beyond the point of maximum productivity.

[4] [Barone] *Principi di economia politica*, Rome, 1913, p. 6.

of cost, supposing implicitly that this is precisely the order in which the firms came on to the market, or else are driven from it, according to whether there is an increase or decrease in demand for the product. He concludes from this that the market price is equal to the cost of production of the 'marginal' firm, and that therefore the supply curve in a competitive market always displays increasing costs.[1] Barone's procedure is formally identical to Ricardo's, in which pieces of land are successively put under cultivation. All Barone does is to substitute the firm for the different pieces of land, efficiency for fertility, and profit for rent. However, such a procedure ignores a fundamental difference: when one wishes to extend cultivation in general (assuming, with sufficient approximation to reality, that the land is used only for agriculture), one can have recourse only to those pieces of land that were not thought suitable before, that is, the worst pieces of land. But if the number of firms in a given industry increases, nothing ensures that the last ones to appear are the least efficient, since they, contrary to the marginal pieces of land, were not unused beforehand, but formed part of another industry. The firms that transfer to the expanding industry are those that could accomplish the transfer with minimal cost, that is, probably, those that were in an allied industry or anyhow possessed capital and labour capable of greater mobility. Equally *vice versa*—if there is a decrease *[299]* in the demand for a given product, the firms which can most easily change their production will leave that industry. Certainly, some firms will be driven out of every industry and will fail, just as in the case of an increased demand entirely new firms will be formed. Barone seems to consider only these cases. But since the firm, much more than being the person of the entrepreneur himself, is formed from a mass of capital and workers, even if a part of the capital is destroyed and a part of the work-force remains unemployed, another part will of necessity be transferred from one industry to another, and it will not always be the most inefficient, but the one that is most easily transferred. To give an example, let us suppose that in industry A, a firm that produces at low costs, has an annual profit of 20, and another firm, that produces at higher costs, has an annual profit of 10. Let us suppose that they foresee that if they changed over to industry B, the first firm would have a profit of 18, and the second, of 5, and therefore in these conditions the changeover does not suit either of them. If, however, the demand for product A decreases, and in consequence, the profits from the first undertaking fall to 15, and those of the second to 6, it is clear that it will be the more efficient firm that is 'expelled' from industry A. The case of the different firms should not be treated analogously with the extension of *all* agriculture to uncultivated land, but rather in a way similar to the extension of cultivation of *one single* agricultural product. In such a case, it is no longer

[1] *Op. cit.*, p. 14. *Cf.* below, p. 42 and p. 42, note 1.

a question of diminishing returns, because the pieces of land on which cultivation will begin will not, usually, be uncultivated lands, but lands already cultivated which, at the new prices, can obtain an increase in rent by changing the kind of cultivation—and they might also be the most fertile ones. The distribution of crops on the different pieces of land is determined not on the basis of the law of diminishing returns, but on the basis of the principle of comparative costs, that is, in an analogous manner to that according to which industries are distributed among different countries.

After this much has been said on the nature of increasing costs, little remains for us to add on the collective supply curve of the industries that are found to be in these circumstances. This is the curve which must represent for every quantity of commodities the price necessary for the production of that quantity by the industry as a whole. As far as *[300]* the construction of this curve is concerned, we can consider the whole industry as a single firm which employs the whole of the 'constant factor',[1] and employs successive doses of the other factors in the amounts necessary to bring production to the required level. For well known reasons, which it is pointless to repeat here, the marginal cost, that in industries with increasing costs is identified with the cost of the unit of commodity produced in the most unfavourable conditions, will be, for every quantity, equal to the price necessary for that quantity normally to be produced. The collective supply curve in conditions of increasing costs therefore represents the marginal costs.

But this procedure, however formally correct it may be, ignores the main problem in the study of an industry in condition of free competition, in which the general equilibrium is the result of the series of individual equilibria which the competing firms must reach independently of one another. To show clearly these relations between the individual and the industrial collectively, it is necessary to reconstruct the passage from the individual supply curve to the collective curve.

The similarity between the demand curve, based on decreasing utility, and the supply curve under increasing costs, based on diminishing returns, is such that one can easily be led to believe that the individual curves are, in both cases, combined by means of an identical procedure. For demand, it is sufficient to add up the quantities which the individual consumers are prepared to buy at a given price to obtain the quantity that, at this price, is demanded by the community. That is to say, the collective demand curve is

[1] *Cf.* Marshall, *op. cit.*, p. 835.

obtained by adding up the individual curves along the abscissas.[1] The collective curve is, therefore, only an enlargement of the individual curve, that is made possible by the fact that the causes of the decrease of the demand price as the quantity of the commodity available increases, have their roots in the nature *[301]* of the individuals, independently (it is supposed) of the fact that there are many or few consumers of that commodity. But this cannot apply to diminishing returns. The cause of this decrease—the fact that one of the factors cannot be increased—operates only for the industry as a whole. The quantity of that factor available for the totality of the producers is constant, but the single producer can increase or decrease the quantity that he uses of it without appreciably influencing the price of the factor itself. In the case of agriculture, 'land from the point of view of the individual cultivator is simply one form of capital'.[2] It is therefore possible that, while the industry has increasing costs, the single cultivator might, up to a certain point, increase his production while lowering his own private cost of production, because he can take advantage of the economies of large scale production, and yet, without being forced to intensify the exploitation of the constant factor, can obtain for himself a large quantity of it at the expense of his competitors. But although this is possible for each producer separately, it is not possible for the totality of producers, and therefore the sum of a series of individual curves of this kind is absurd, since each one of them is valid only on condition that the production of the other individuals remain unchanged. In order to make it possible to add up the individual curves it is necessary to have recourse to a stratagem that moves the cause of the increase in cost from the conditions of the industry to the conditions of the single producer. This is achieved by supposing that the number of producers is fixed, and that each of them, with the increase in his production, cannot increase the quantity used by him of the factor of which there exists a fixed quantity for industry as a whole, so that the individual cost of production has to increase. In these conditions, the individuality of the 'enterprise' is no longer characterised solely by the unity of management, that is, by the entrepreneur, but also by the presence of a unit of the 'constant' factor. In this way the formation of the collective supply curve, by means of the addition of the individual curves, becomes possible. *[302]*

---

[1]Strictly speaking, the individual demand curves also need a further hypothesis, if they are to be added. It must be supposed that every purchaser wants to buy only what he can consume, excluding the possibility of reselling the commodity bought. Otherwise, at prices lower than the market price, everyone would be prepared to buy an unlimited quantity of the commodity.

[2]Marshall, *op. cit.*, p. 170.

III. *Decreasing Costs*

The principle of decreasing costs has arisen as a generalisation of the commodity observed fact that the cost per unit product for a firm, decreases with an increase in the quantity of the commodity produced *by that firm*. Such a decrease derives essentially from two groups of causal elements. A first group relates to the possibility of having recourse to better methods of production when the size of the firm increases. This is the possibility of introducing 'internal economies' (of which a characteristic and principal element is a greater division of labour). We will not dwell on this case except to mention that it is distinct from that previously discussed (p. 9) of increasing productivity of a variable factor of production applied to another that remains constant. In that case the more than proportional increase of the output is due solely to the fact that initially one is forced to use an excessive quantity of one of the factors (the constant one), which hence has a negative effect on the output. (Which is to say that initially the product turns out to be less than what it would be if it were possible to use a smaller quantity of the constant factor.) With the increase of one of the factors of production, the *proportion* in which they are combined becomes more favourable. In the case we are considering here what is essential, on the contrary, is the variation of the absolute *size* of the totality of the factors used; whilst it is possible that the proportion between them does not vary.[1] *[303]*

The first group of causal elements determines in the first instance a tendency towards a decrease of the *marginal* cost; and it only through such an effect that it leads to the decrease in the average cost of production.

---

[1]It must be recognised that, from the point of view of the causes that determine the decrease in cost, the distinction allows a number of intermediate cases to exist. In the case of a single firm considered here, it is possible that, if it is very small, the minimum quantity that it can use of a given factor is relatively so large that it has a negative productivity. On the other hand, the impossibility of using a smaller quantity of the constant factor from which the initial lesser productivity derives, is often identified with the impossibility of using, in those conditions, better methods of production. The distinction does not, because of this, lose its *raison d'etre*. The first form, based on the proportion of factors, is characteristic of the totality of all the industries that use a given factor of production, while for one firm amongst many under conditions of competition, it is generally possible to procure the different factors in such a way as to combine them in the most propitious proportion. The second form, based on the size of the totality of factors used, is relevant only in the case of one firm. While in the case of a group of industries, the limit of size below which production is less efficient is generally exceeded. In other words, the two cases apply to different orders of magnitude: the firm and the industry or, better, the group of industries.

The second group of causal elements derives from the fact that every firm must bear a certain quantity of 'overheads' which, with the increase in production by the firm, remain constant, or, at least, increase less than proportionally. From the possibility of distributing such overheads over a larger number of units produced, there results a tendency towards a decrease of cost of each unit. It is therefore clear that these elements can only precipitate a decrease in *average* cost of production, while they do not exert any influence on the marginal cost. Marginal cost would, to a certain degree, be presumed to be increasing, without this counter-balancing the effect that the decrease in overheads allotted to each unit has on average cost. This case apparently present a closer analogy with decreasing costs derived from variation in the proportion between the factors of production. It might appear to be correct to consider overheads as 'constant factor', and particular expenses as the 'variable factor' which is applied to the first in successive doses, and then to infer an analogy in the elements that in both cases determine the decrease of cost with the growth of production. But in reality there is a profound difference: what decreases in the case of 'overheads' is only the average cost, while in the case of the 'constant factor' (as in the case of 'internal economies') the essential thing is that the marginal cost decreases and average cost only decreases too as an indirect effect.

The cases in which productivity grows as a consequence of variations in the size of the single firm cannot be accommodated in the theory of price determination in a regime of free competition, since it is clear that, if a firm can decrease *[304]* its costs without limit by increasing production, it would continue to reduce the selling price until it had acquired the whole market. We would then have abandoned the hypothesis of competition. We will, therefore, not stop to analyse such cases. They cannot, however, be totally ignored, for many writers consider them to be the principal basis of the tendency towards decreasing costs in a regime of competition. Cournot[1] believed he could form a collective curve of decreasing costs under competitive conditions, simply by adding up individual curves representing the decrease of the cost per unit for each producer from the increase of his individual production; without noticing, as Marshall noted, that such premises 'lead inevitably to the conclusion that, whatever firm first gets a good start will obtain a monopoly of the whole business of its trade in its district'.[2] Even

[1][Cournot] *Recherches sur les principes mathematiques de la theorie des richesses*, 1838, para. 48, p. 96 *et seq.*

[2][A. Marshall] *Principles*, p. 459, note.

Edgeworth fell into an error of this sort[1] but he rectified it[2] following the publication of Marshall's book, which clarified the question in a definitive manner and remove all possibility of doubt. Barone, however, persisted in the belief that the mistake had not been rectified, even after Marshall's publication. He denies the possibility of a static curve under conditions of decreasing costs when there are several competing firms: 'the decreasing curve can have a concrete and precise meaning in case (a) (a firm considered in isolation) and in case (b) (a monopoly); but in case (c) that is, of several competing firms, we cannot succeed in understanding what it means'.[3] Barone obviously thought that the erroneous method followed by Cournot was the only one imaginable by which a collective curve could be formed under conditions of decreasing costs, forgetting that the theory of 'external economies' allows for the perfectly correct construction of such a curve, at least in a formal sense. *[305]*

But the reason why this form of increasing returns especially interests us, is the part that it has played—together with decreasing returns due to variations in factor proportions—in the genesis of the theory of the equilibrium price of individual commodities, and the considerable influence that it still exerts in making this same theory acceptable.

Marshall has played such a predominant role in the formation of this theory that it is sufficient, for the purposes of our investigation, to limit ourselves to a consideration of the evolution of his thought. In *Economics of Industry*,[4] which contains the first complete expression of his doctrine, Marshall makes the law of increasing returns derive directly from the 'Law of Division of Labour' (p. 57), and considers this to be dependent in the first place 'on the size of the factories in which the work is done' (p. 52); thus in assuming among the causes for the decrease of cost a condition which is compatible with free competition, he skims over the error that later he was himself to correct. In the second place, many of the advantages of the division of labour 'can be secured by small factories and workshops, provided there are a very

---

[1][Edgeworth] 'On the applications of mathematics to political economy', *Journal of the Royal Statistical Society*, 1889, pp. 570 - 571.

[2][Edgeworth] *Papers Relating to Political Economy*, London, 1925, Vol. ii, pp. 305 - 306, note.

[3]*Op. cit.*, p. 197, note.

[4][A. Marshall, *Economics of Industry*] London, 1879; the 2nd edition, from which we quote, is from 1881, that is, it precedes by ten years the *Principles of Economics*.

great number of them in the same trade' (p. 52). Marshall perceives these latter advantages as being mainly in the development of subsidiary industries, those that make the tools and the machines necessary for production in the industry under consideration and which facilitate inter-relations between the different branches of the industry. But, he immediately warns, small factories can make use of these advantages only if 'many of them are collected together in the same district' (p. 53). The *localisation of industry* is therefore a necessary condition for verifying this form of increasing returns.

As can be seen, in this formulation those circumstances which were later to be considered the fundamental cause for the decrease of cost, that is, 'external economies', are found only in embryo and as secondary elements. The fact that their influence was conditioned by the localisation of industry makes it apparent that they could not be at the root of the *[306]* tendency towards increasing returns connected exclusively with the increase of production. We cannot in general presume that to every increase of production there corresponds a greater localisation of industry, and to every decrease a spreading of factories over a wider territory—a presumption that would be necessary to establish the dependence of decreasing costs on economies stemming from the localisation of industry.

And as to the other sort of external economies, namely those improvements in the methods of production which follow an increase in the size of the industry, Marshall rejected the idea that the decreases in the cost deriving from such improvements could be considered exclusively as an *effect* of the increase of production, pointing out that 'the general progress of knowledge would in any case have done much towards bringing about such changes' (p. 92)—an observation which seems to us to carry weight, although it was later ignored by Marshall himself.

But when he noticed that a decrease of cost, deriving from the increase in the size of the factories and from a larger division of labour, was incompatible with free competition, he abandoned his original point of view, and instead expanded his theory of external economies, to the extent of considering these as the sole cause of decreasing costs in a regime of competition.

It is only in the *Principles of Economics* that the theory appeared in its definitive form. The radical change that this work precipitated in the substance of the laws of variation of costs went largely unnoticed, while the theory of value based on the 'fundamental symmetry' of the forces of demand and supply, of which those laws are necessary premises, remained unchanged. In essence, the foundations were replaced without the building standing above

receiving a single jolt from it all, and it was the great ability of Marshall which allowed the transformation to pass unnoticed. If he had given the originality of the new conception the prominence it deserved, perhaps it would not have been received without opposition. By presenting it as something very well-known and lacking novelty, almost as a commonplace, he was able to have it accepted as a tacit compromise between the necessities of the theory of competition, which are *[307]* incompatible with the decrease of individual cost, and the necessity not to stray too far from reality that (being far from perfect competition) presents numerous cases of individual decreasing costs of this kind. The fact that the 'external economies' peculiar to an industry, which make possible the desired conciliation between scientific abstraction and reality, are themselves a purely hypothetical and unreal construction, is something that is often ignored.

The characteristics of the new theory are clarified in the process of forming the collective supply curve under a regime of competition. The external economies constitute a link that unites the conditions of production of the individual firms in the industry. The cost of production of each firm is not determined solely by the quantity that it produces itself, but also, at the same time, by the quantity produced by all the other firms. In studying the individual equilibrium, three variables must therefore be considered: cost, quantity produced by the single firm, and quantity produced by the industry as a whole.

The hypothesis of free competition fixes the limits between which the theory of decreasing costs based on external economies is applicable. It implies that, by considering 'an industry' as the set of firms that produce a given commodity, each firm must be so small relative to the industry, that the influence of a variation in the quantity produced by the firm on the market price can be taken as negligible. Further, supposing that each factor of production is used by a large number of different industries, a variation in the quantity of it used by an industry does not exercise any appreciable influence on the remuneration of that factor, since this is determined by the general conditions in the totality of the industries that use it.[1] The quantity of factors

[1]Pigou explicitly states that such a procedure is destined to be applied to 'a great number of different industries and occupations, each one of which is supposed to make use of only a small part of the aggregate resources of the country. Because every occupation is thus relatively small, the price per unit of the several factors of production in each occupation is determined by the general market conditions, and is not effected to any appreciable extent by variations in the quantity of them that is employed in that occupation', *Economics of Welfare*, p. 935. And also *cf.* Bowley, *Mathematical Groundwork of Economics*, Oxford, 1924, p. 28.

that each industry can obtain *[308]* for itself at the market price must thus be considered as practically unlimited.

Let us now begin by examining the shape of the supply curve of a single representative firm. We represent on the abscissa (see Fig. 2) the quantities of the commodities produced by the firm, and on the ordinate the corresponding unit costs, that is the total cost of each quantity divided by the number of units produced. To satisfy the above conditions, such a curve must of necessity conform to a well-defined type. First of all, it cannot display increasing costs for all of its length: because in such a case competition would tend to make every firm infinitely small and the number of firms infinitely large. Hence, because of the need for each firm to reduce its own production

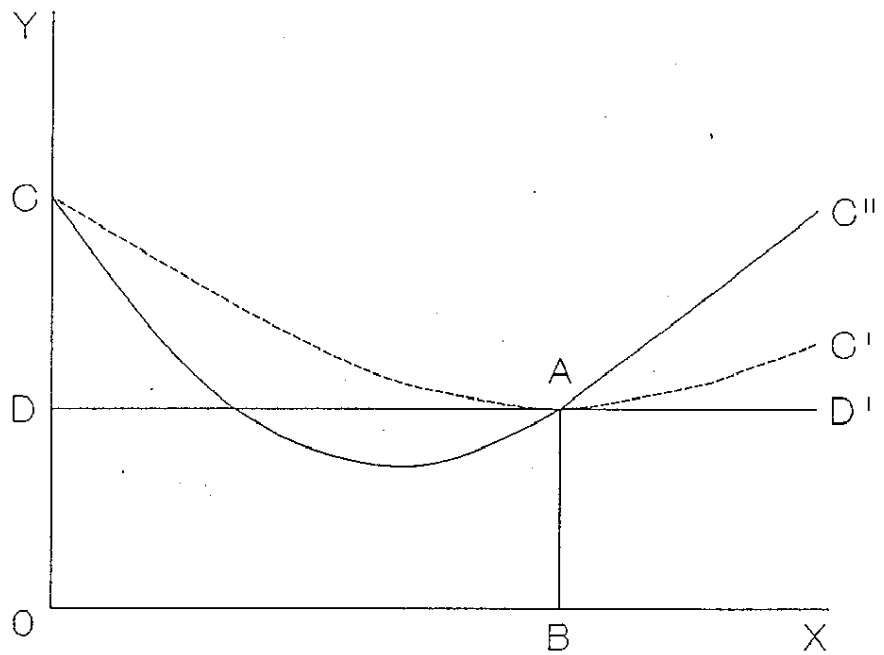

Figure 2

so as to reduce its costs, there would be no possibility of achieving any equilibrium *[309]* whatsoever. The curve must therefore, in each case, initially display decreasing costs. Secondly, it will not show solely decreasing costs, since if it did, a firm would necessarily acquire a monopoly in the industry, contrary to the hypothesis of competition. The supply curve of the representative firm will therefore have in each case a shape of the type CC'.

Given the form of the curve, it will possess a *minimum*,[1] which corresponds to the point of maximum economy, that is of the quantity that it can produce at least cost. This curve presupposes, among its conditions, that the industry as a whole produces a fixed quantity, let us say *z*. With the variation of this quantity, the form of the individual curve may be modified, since it is supposed that the conditions of production of the individual firms that compose the industry are not independent of one another. To a collective industrial production equal to *z* will correspond a specific general equilibrium price, which will also be the only possible selling price for the firm in question. The regime of competition is defined as a state of affairs in which each 'bears the market price without trying deliberately to modify it' which means he can 'suppose that the price is constant'.[2] That means that from the point of view of the single producer, the market demand curve is a straight line parallel to the abscissa. This is the only way of representing a state of affairs in which a producer can sell a practically unlimited quantity at the market price. The demand curve (DD'), understood in this sense, will always be tangential to the individual supply curve (CC') at the point of maximum economy (A). That is, each firm will always sell at the minimum unit cost of production. In fact the CC' has been traced in such a way that its ordinates represent the total remuneration of *all* factors of production used, including the "organisational" factor.[3] Now if the supply curve *[310]* had at some point ordinates less than AB and therefore intersected the demand curve, that would mean the firm considered would have the possibility of producing at a unit cost lower than the market price, and thus of obtaining a higher than normal profit. But whatever it was that gave rise to such superiority would itself be a factor of production, and the supposed abnormal profit would be nothing other than the remuneration of this factor, which would have been arbitrarily excluded

[1]Exceptionally the curve can show different *minima*; in such a case the *minimum minimorum* should be considered.

[2]Pareto, *Cours d'economie politique*, para 46 and note.

[3]It is useful to reiterate that this relates only to conditions of perfect competition, that is to a situation similar to the 'etat limite' of Pareto, 'characterised by Walras' hypothesis of an ideal manager who realises neither gain nor loss (his salary as the director of the firm being included in the expenses of production), *Cours*, para 87. The criticisms of Walras's and Pareto's views raised by Edgeworth ('On the Use of the Differential Calculus in Economics', *Scientia*, 1910, I, p. 92, *et seq.*) can show that this state of affairs is not typical, but they do not prove that, within the given hypothesis, the conclusion is not correct.

from the list of elements constituting the cost of production.[1] Therefore, when we take into account all the costs borne, the total revenue of each firm will exactly balance the total expenses.

This conceptualisation must be used with caution, so as not to fall into the vicious circle of including among the costs, that is among the conditions that contribute to the determination of the price of the product, quantities that are determined by that price and vary with it.[2] Thus, if a factor of production of which there exists a constant quantity were used only, or predominantly, in the industry considered, its remuneration would be the effect, and not the cause, of the price of that particular product. It would, therefore, not be a part of the cost of production, but 'surplus' or 'rent'. In reality the conditions that give rise to abnormal profits of the sort indicated (e.g. favourable position or exceptional managerial ability, goodwill etc.) generally fall into such a category and cannot properly be a part of the production cost. *[311]* But this happens precisely because these conditions are outside the limits of the conditions that we have taken as being characteristic of free competition (p. 27). When it is supposed that all the factors of production are used by a large number of industries (and thus also that they are completely transferable from one to another) their remuneration, from the point of view of each of the industries, is fixed, and cannot, from such a particular point of view, be considered as rent.[3]

The individual supply curve under a regime of competition, also presents another peculiarity. If we call the marginal cost of a firm the difference between the total cost that it must bear to produce a quantity x (when it is organised to produce x), and the total cost for it to produce $x + x$ (when it is

---

[1]It is almost superfluous to add that, in the opposite case, that is if all the ordinates of the supply curve were greater than AB, the firm considered would not be able to sell anything at the market price and therefore would be eliminated from the industry.

[2]Marshall senses the danger of this vicious circle into which it is easy to fall when we approach the actual conditions of 'the world in which we live'. 'Present incomes earned by them [by the appliances of production] will be governed by the general relations between the demand for and the supply of, their products; and their values will be arrived at by capitalising these incomes. And therefore, when making out a list of normal supply prices, which, in conjunction with the list of normal demand prices is to determine the equilibrium position of normal value, we cannot take for granted the values of these appliances for production without reasoning in a circle'. *Principles*, p. 810, and for a concrete example, *cf.* p. 417, note.

[3]In favour of this supposition and of its application in the particular case, see Pigou, *op. cit.*, p. 933, note.

organised to produce x + x) we can deduce from the average cost curve, a curve that represents the variations of marginal costs (CC" in Fig. 2). Such a curve is constructed by analogy with the marginal productivity curve examined on p. 6. The marginal cost curve will in each case intersect the average cost curve at the point of maximum economy (A) which is also the only possible point of equilibrium.[1] Which means to say that the average cost and marginal cost of each firm in every state of equilibrium will always, under the stated assumptions, be identical.[2] By producing a quantity OB and selling at the price AB the firm will simply *[312]* receive reimbursement of expenses, without any producer's rent being left over.

A perfectly possible case is that in which the individual marginal cost is, for some or even for all the amounts of production, constant. For such amounts the marginal cost curve would coincide with the average cost curve, and within these limits the equilibrium will be indeterminate, given the definition of competition that we have followed up to now. Such an indeterminacy can be eliminated, if there is added to this definition of competition the attribute that Pigou considers to be fundamental (and which does not contradict the definition we have adopted) when he defines 'simple competition' as a set of 'conditions under which it is to the interest of each seller to produce as much

---

[1] *Cf.* above, p. 7—the analytical demonstration of this property is given by Edgeworth ('Railway Rates, IV', *Economic Journal*, 1913, p. 214), who, however, interprets the pair of curves in a different manner from the one followed here. We have only briefly indicated the general relations between the average cost curve and the marginal cost curve, that have been made well- known by the treatment of the subject by Pigou, *Economics of Welfare*, 1920, app. III.

[2] Such equality, which is generally ignored, has been pointed out by Flux, 'Just at the turning point from decreasing to increasing costs, that is, at the point of maximum economy, the marginal expenses per unit become identical with those average expenses which cover prime costs and supplemental costs together, and represent the proportion of total expenses to total output' (*Economic Principles*, 2nd edition, London, 1923, pp. 61 - 62). Given this equality, the following question put forward by Pantaleoni does not appear to be admissible, 'Why does the price in firms that work under increasing costs and in conditions of free competition, tend to balance the marginal cost, and, on the contrary, in firms that work under decreasing costs tend to balance unit cost' (*Temi, Tesi, problemi e quesiti*, Bari, 1923, p. 82, note 255). We note that when J. A. Hobson, concluding one of his polemics against the 'marginalists', writes 'In other words, the so-called final or marginal productivity turns out to be nothing other than an average productivity....The whole notion that there is a marginal increment...is entirely fallacious' (*The Industrial State*, 2nd edition, London, 1910, p. 116); he cannot from our point of view be considered entirely wrong (as Marshall, however, was to declare, *Principles*, p. 517, note); his statement is wrong only in the second part, *precisely because* it is correct in the first part.

as he can at the ruling market price'.[1] Under these conditions, if the unit cost curve displays constant costs over a certain range, equilibrium will be reached at the point which corresponds to the maximum quantity that can be produced at that cost: and it will no longer be possible to allow that the curve may display constant costs throughout its length, for this would lead to monopoly on the part of the firm considered.

It has been said above that, because the link of external economies exists between the conditions of production of the different firms, the pair of individual curves represents the conditions of a single firm only in a given state of the industry; for example, when the quantity produced collectively is z. In the absence of external economies the individual curves would remain unchanged with variation of z. The increase of collective production would derive from an increased number of firms, while each of these firms would continue to produce the same quantity at the same cost. The collective supply curve would display constant costs, the cost of the factors of production being taken as constant. *[313]*

But, given the presence of external economies, the form of the individual curves would be completely altered with the growth of z. The point of maximum economy could be moved in any direction because of the change, corresponding to larger or smaller individual outputs. But in every case the lowest individual cost should decrease with the increase of the quantity produced collectively. Under these conditions, the collective supply curve must be formed in the following manner. Since each individual curve shows, in general, only one point of possible stable equilibrium for each quantity produced collectively, only these points would figure in the composition of the collective curve. All the others (in Fig. 2 the descending and ascending parts of CC") represent conditions that would be realised only with the failure of the assumed perfect competition, e.g. in the period of time necessary to pass from one equilibrium to another. On the basis of Fig. 2 we will imagine a third axis OZ, normal to the plane of the paper and passing through O, on which are measured the quantities produced by the industry as a whole. For each value of z, we will get a different pair of curves, that will give rise to two surfaces which will intersect in a curve with three co-ordinates. This will represent the locus of the point of maximum economy for the individual firm.

This new curve represents the variation of individual costs as a function both of the quantity produced by the firm considered, and of the quantity produced by the totality of firms. For each of the firms there is a curve of this

[1] [Pigou] *Op. cit.*, p. 190.

type; not only for firms that exist in a given state of the industry, but also for those that will enter to form a part of it when the quantity produced collectively is increased. By arranging all these individual curves along the x axis (that is, by adding the quantities produced by the individual firms) we get a flat curve, the collective supply curve, since the sum of all the individual values of x, corresponding to a given value of z, is equal to this value of z.

Such a curve represents collective average costs, that are, for every quantity produced, equal to individual average costs, and therefore also to the individual marginal costs which coincide with them in *[314]* every state of equilibrium. For each quantity of the commodity, these collective average costs are equal to the price that it is necessary to pay so that the industry can continue to produce that quantity. In fact, having paid the average cost, all the factors used are remunerated at the current price, and no residue remains. It is thus the curve of collective average costs that, with the demand curve, contributes to the determination of the price of the commodity.

With a procedure analogous to that followed in the case of individual curves, we can, from the collective average cost curve, deduce a corresponding marginal cost curve. This curve has no direct relevance to the determination of the price in conditions of competition, and it is therefore outside the scope of our argument. However, we mention it because it characterises the nature of external economies. In conditions of decreasing average costs the collective marginal cost is, for any quantity produced, less than the collective average cost; and since the individual marginal cost is in any instance equal to the latter, the result is that collective marginal cost is less than the corresponding individual marginal cost. The reason for this divergence lies in the fact that, when calculating the individual marginal cost, we take into account only that part of the increase of output resulting from an increase in the expenses of a single producer, and that he is able *to appropriate for himself*. But when the expansion of the industry leads to greater external economies, the single producer cannot appropriate all the increase of output derived from the increase of his expenses, since, as all the producers of that industry have the possibility of availing themselves of the new external economies, their output too will be augmented for a constant level of expenses (even if, in this case, by an infinitesimal amount).[1] Now in calculating the collective marginal cost,

[1]For simplicity of exposition we have ignored the fact that, in order to have a noticeable effect in the form of external economies, the increase of output of the order of magnitude of the individual increment of production of one among many competitors is not sufficient. It must be noticeable, even if small, compared with the size of the totality of the industry. The (continued...)

we take into account these benefits that the performance of each producer brings to all the others, without the latter having had any influence. This is the reason why the collective marginal cost *[315]* is less than the individual cost. We note, incidentally, that one of the proofs of the impossibility to realising maximum collective utility in a system of perfect competition is based on this divergence. Each producer has an interest in only taking his production up to the point at which the increment of output he obtains equals in value the increment of his expenses. But it is not rewarding for him to take output beyond this point, even if the loss that he would suffer is less than the advantage that he collectively would obtain. In other words, under a regime of competition, equilibrium is achieved at the quantity of output that equates the demand price and the collective average cost, while maximum utility is obtained at the quantity at which the demand price and the collective marginal cost are equal.[1] *[316]*

IV. *Constant Costs*

We have up to this point considered separately the causes that tend to make cost increase with the increase of production, and the causes that tend to make it decrease. But, strictly speaking, there is no logical difficulty in supposing that the two groups of causal elements can operate simultaneously. Thus, it is possible that in an industry which uses the totality of the existing quantity of a factor of production, and therefore has a tendency towards increasing costs, the increase of production carries with it an increase in external economies, such as to give rise to an opposite tendency. The two

---

[1](...continued)
effect considered, therefore, will occur only if a certain number of firms increase their production at the same time.

[1]This doctrine is substantially due to Marshall (*Principles*, Book V, Chapter XII), but the mention of it here follows the lines of Pigou's deeper and more precise analysis, (*Economics of Welfare*, 1920, app. III, modified in part in the second edition, 1924, especially, p. 194). The observation that, in the case of decreasing costs, the point of equilibrium cannot be found on the collective marginal cost curve, but must correspond to the average cost, was made by Commons (*The Distribution of Wealth*, New York, 1893, pp. 125 - 126), who did not go so far, however, as to extent this concept to the point of identifying in the collective average cost curve the locus of equilibrium points, that is, the true collective supply curve. The 'dual system' of collective supply curves was put forward for the first time by Pigou ('Producers' and Consumers' Surplus', *Economic Journal*, 1920) and modified in his subsequent writings. Edgeworth, who at an earlier date (review in *Economic Journal*, 1894, p. 686) had rejected Common's statement, later accepted its guiding principle and contributed greatly to perfecting that theory.

tendencies will in part compensate each other and a diminished variability of costs will result. In the case in which the two opposing forces are equal, they will cancel each other out, and the cost will remain constant with variation of the quantity produced. This latter case is certainly exceptional, but it would be arbitrary to infer that industries with constant costs occur only exceptionally. It can be supposed, much more simply, that it is not the canceling out of the two opposite tendencies but the absence of both, that gives rise to the case of constant costs. If all the factors of production used by an industry are used in many others and if the conditions of production of the individual firms are independent of each other, the industry operates under conditions of constant costs. These assumptions are not improbable. On the contrary, the remote probability of assumptions that give rise both to one and the other tendency towards variability of cost, would seem to indicate that the absence of both is to be considered a more general case—given the conditions of particular equilibria—than the presence of one of them. Therefore the case of constant costs, rather than those of increasing and decreasing costs, should be regarded as normal. This must have been Ricardo's opinion, since he states that commodities which can be produced at constant costs constitute 'by far the greatest part of the goods that are daily exchanged on the market'.[1] *[317]*

But, as we said above, the theory based on the symmetry between the forces of supply and those of demand holds good only on condition that the variability of the cost of production with the variation in the quantity produced has the same degree of importance as the variability in the demand price. The greater the importance of cases of constant costs, the greater the influence of cost of production in determining the price, the greater the disturbance to that symmetry. This is probably the explanation of the otherwise surprising fact, that all the writers who hold that theory take only the most complicated and unlikely form of constant costs into consideration, ignoring the most simple and obvious one. Thus, besides Marshall, we find Sidgwick writing that constant costs 'can *only* result from the accidental balance of two opposite tendencies',[2] and similarly Palgrave's *Dictionary*: 'In general, the increase of the scale on which an industry is carried on is accompanied by a change in the proportionate cost of its product; but when the increased difficulties of extractive industry...are set off against the economies arising from improved organisation in manufacture, we may find an exact balance struck, and an increased produce obtained by labour and

[1][D. Ricardo] *Principles*, in *Works*, p. 10. *Cf.* also J. S. Mill, *Principles*, Vol. I, p. 547.

[2][Sidgwick] *Principles of Political Economy*, 1883, p. 207.

sacrifice increased just in proportion. In such a case the law of *constant returns* is said to hold.'[1] Finally, someone has taken this point of view to its logical conclusion, and is led to argue for the quasi-impossibility of constant costs: 'In current discussion it is usually assumed that there will be many cases in which the marginal cost will remain stationary as the output of an industry is increased, so that we may have a law of constant cost. But such a result could be brought about only by an accidental equivalence of the various contending forces which are set in operation by an increased demand for any commodity. In almost all cases the chances would be greatly against a precise balance of these opposing influences, so that, in strictness, we must *[318]* conclude that the usual result of enlarging the output is to raise or lower the marginal cost.'[2]

It has been noted that 'to treat *variables* as *constants* is the characteristic vice of the unmathematical economist';[3] and others have added that, of this vice 'a striking and important instance is to be found in the treatment of cost of production as a constant, and the consequent failure to recognise the part played by demand in the determination of normal, as well as market, value.'[4] We must ask ourselves if, in the case we are considering, the mathematical economists have not gone too far in correcting this vice, so much so, as to fall into the opposite vice, that is, treating a constant as a variable. *[319]*

---

[1][*Palgrave's Dictionary of Political Economy*] Vol. II, p. 582.

[2]C. J. Bullock, 'The variation of productive forces', *Quarterly Journal of Economics*, Vol. XVI, p. 500, *cf.* note.

[3]Edgeworth, *Mathematical Psychics*, London, 1882, p. 127, note.

[4]J. N. Keynes, *Scope and Method of Political Economy*, p. 263.

V. *Co-ordination and Critique of the Three Tendencies*

Having examined separately the hypothetical conditions that give rise to tendencies towards increasing, decreasing or constant costs, respectively, it is necessary to consider them in their entirety, so as to understand whether, and within what limits, a co-ordination of the different tendencies under one single 'law of non-proportional costs' is admissible; bearing in mind that the aim is to arrive at a general and organic conception of the supply curve, such that ultimately this curve is symmetrical to the corresponding demand curve for each commodity.

The first difficulty which inhibits this co-ordination derives from the fact that the hypotheses on which the different tendencies are based were originally, as we have noted, designed with different objectives in mind. The hypotheses of diminishing returns—according to which we take a given factor of production and isolate the conditions that are essential to the determination of its return, are appropriate to the study of questions of distribution. The hypotheses of increasing returns according to which the price of factors is fixed by external elements, and which concentrate attention on specific commodities, are suitable for the study of the conditions that influence the price and quantity produced of individual commodities. The hypotheses of diminishing returns were connected originally with the theory of rent, that is with the first case identified of marginal distribution of the product among the factors. Ricardo used them to investigate, not the laws that regulate the price of the product, but rather the laws of rent, and his *Principles*, of which these hypotheses are characteristic, is essentially a treatise on distribution. For Ricardo and his contemporaries, 'to determine the laws which regulate . . . distribution, is the principle problem in Political Economy'.[1] Modern economists, however, are generally oriented towards the problem of determining the prices of individual commodities (so much so that they have included within this the study of distribution, considered as the determination of the *[320]* prices of the factors of production). It is from this new point of view that the characteristic hypotheses of decreasing costs arose. The analysis based on such hypotheses 'is not designed for application to the output of a whole body of a country's resources lumped together into a single industry. Its purpose, on the contrary, is to provide machinery for studying the distribution of resources among a greater number of different industries and occupations, each one of which is

[1] [D. Ricardo] Preface to the *Principles of Political Economy*; see also *Ricardo's Letters to Malthus*, Bonar edition, Oxford, 1887, p. 175.

supposed to make use of only a small part of the aggregate resources of the country'.[1]

Therefore, the two groups of hypotheses in question, rather than referring to different phenomenon, represent different aspects under which the same phenomenon can be considered. That is to say that the applicability of one or of the other group depends, in many cases, not so much on the objective conditions of the economic system studies, as on the nature of the problems that we propose to study in respect to it. The element of arbitrariness that is thus introduced into the criterion that should guide us in a classification of industries according to the manner of the variation of cost, is evident in the choice of the characteristic that is to be taken as the basis of a definition of 'industry'. If every single industry is defined as the exclusive consumer of a given *factor of production* (for example, agriculture, the iron industry, etc.), a condition is at once assumed that implies a tendency towards increasing costs for the industry, since it is precisely the factor that is characteristic of the industry (cultivable land, iron mines, etc.) that, with the increase of production, generally remains constant. If, on the contrary, every industry is defined as the sole producer of a given *product*, and this is meant in a fairly restrictive sense, so that in general it can be thought that every industry uses only a small fraction of each factor of production (negligible in comparison to the quantity used by all the other industries together), we thereby exclude from the industry the circumstances that generates increasing costs and make it more probably subject to the law of constant costs, or, in further specific conditions, to the law of decreasing costs.[2] This *[321]* derives from the fact that, as we have seen, increasing costs are the result of variations in the *proportion* between the factors of production used, while decreasing costs derive from variations in the *absolute quantity* of the totality of factors.

---

[1]Pigou, *op. cit.*, p. 935.

[2]We have used the expressions 'diminishing (or increasing) returns' and 'increasing (or decreasing) costs' as equivalents. However, to give greater prominence to the contrast hinted at in the text, we have preferred, when this would not have confused the reader, the first form. This form refers to a quality of factors (productivity) in the case of decreasing productivity; and the second refers to an attribute of the product (cost) in the case of decreasing costs. Bullock, who in the article quoted ('The Variation of Productive Forces') has stressed that the forces generating the two tendencies are of different order, proposed a change in terminology, according to which the expression 'economies in organisation' should be substituted for 'increasing returns' so as to avoid this last term being linked with decreasing productivity (p. 489).

Although we have limited ourselves to the consideration of static conditions, one should note incidentally that when in a subsequent approximation the element of time is introduced, this further increases the uncertainty over the classification of industries according to variability of cost. For short periods, conditions generally prevail that approach those of diminishing returns, for given the limited mobility of certain forms of capital and labour, these may be considered to be incapable of being increased unless a long enough time is allowed for the necessary transformations. Whilst, with the increase of the period of time allowed, we move away from such conditions and approach those conditions appropriate to decreasing costs. Thus the same industry can belong to one or the other category according to the length of the period considered.[1]

The heterogeneity between the two groups of hypotheses cannot, in every case, be considered an unsurmountable obstacle to the co- ordination of the two tendencies respectively that originate from them. However, the arbitrary and inharmonious characteristics which vitiates the theoretical system at its starting point, and its inadequacy in clarifying the nature of the operative elements, cannot help but make it less fruitful as an instrument for the study of problems in which only the effects of these causal elements are considered.[2] *[322]*

But the most serious imperfections of the 'symmetrical theory' are inherent in the very nature of these hypotheses, even when considered separately. Let us go back to the conditions which a supply curve of the type that is used in studying the 'particular equilibria' of individual industries must satisfy. Since it represents only two variables, it is necessary to suppose that all the other conditions of the problem remain unchanged with the variation in the production of the commodity. It is necessary, in particular, that the demand of consumers, and the conditions in which other commodities are produced, should not change. That is to say (1) the supply curve must be independent, both of the corresponding demand curve, and also of the supply curves of all

---

[1]Marshall himself has shown 'the unsatisfactory character of these results, partly due to the imperfections of our analytical methods'. *Principles*, p. 809 and *passim*.

[2]It may further be noted, in connection with the heterogeneity of the different tendencies, that the collective supply curve under conditions of increasing costs, indicates *marginal* costs. The curve under conditions of decreasing costs indicates *average* costs, and the curve under conditions of constant costs indicates *average and marginal* costs, (which in such a case coincide). Were a supply curve in part ascending, and in part descending, it would represent marginal costs in the ascending part, and average costs in the descending part. The result is hardly 'elegant'; but, given the premises, it is inevitable.

the other commodities; (2) the supply curve is valid only for slight variations in the quantity produced, and, if we depart too far from the initial equilibrium position, it may become necessary to construct an entirely new curve,[1] since a large variation would, in general, be incompatible with the condition *ceteris paribus*. *[323]*

These conditions reduce to a minimum the range over which hypotheses of increasing costs are applicable to the supply curve of a product. They are satisfied only in those exceptional cases where the totality of a factor is used in the production of a single commodity. But, in general, each factor is used by a number of industries that produce different commodities, and in this case only a supply curve of the totality of those commodities is possible, based on the assumption that the group of industries that have a common factor can be regarded as one single industry, according to the method we have followed above (p. 21). But the supply curve which displays increasing costs for one of the commodities is inadmissible. Let us examine two possibilities: the one appropriate to the case in which we are dealing with a small number of commodities, and the other to a case dealing with a large number of commodities. In the first case, if one of the industries increases its production,

---

[1]Marshall has repeatedly emphasised the importance of this limitation: 'the ordinary demand and supply curves have no practical value except in the immediate neighbourhood of the point of equilibrium', (*Principles*, p. 384, note). Marshall's proposition is important not only because it excludes large variations in the quantity produced, but also because it allows small variations. If the supply curve is to be considered one of the elements that determine price, it is not sufficient that *only the point of equilibrium* is significant. At least those points in the immediate vicinity must *also* be significant; since these represent precisely the forces that will be set in motion when an accidental shift of the equilibrium position occurs, and that would tend to re-establish that position. That is, they are necessary conditions of that equilibrium. It is interesting to point out here how, Ricci, in order to defend Marshall's supply curves for products (at variable costs) from some of the criticisms noted earlier, inadvertently had to abandon precisely that condition necessary for them to have meaning. He in fact writes that the curves of supply 'exist only in relation to a particular and determinate equilibrium. They cannot be used to represent an equilibrium different from the first. Their ordinates, in short, do not say what will be the prices or marginal costs where production amounts to exactly the quantity indicated by the respective abscissas, but say only which costs must be ascribed to successive doses of the quantity produced in that single and determinate equilibrium to which they refer.' ('Curve piane di offerta dei prodotti', *Giornale degli economisti*, 1906, Vol. II, p. 224). Curves thus characterised are not true supply curves, which can enter into the determination of the price of the product. They are, in Marshall's terminology, *particular expenses curves*, destined for very different uses and in which only *'for convenience* the owners of differential advantages may be arranged in descending order from left to right'. Marshall has made use beware of the frequent error of attributing to supply curves the characteristics of particular expenses curves (*Principles*, pp. 810 - 811). Wicksteed's criticisms originated from a confusion of this sort.

it must use a larger quantity of the common factor at the expense of the other industries of the group, so that that factor must be utilised more intensively (that is, combined with a larger proportion of the other factors), and thus, as we know, the cost will rise. But it will rise not only in the industry that has increased production, but also in the other industries of the group; and in each case the increase of cost will be proportional to the degree in which the common factor enters the cost of each, for this common factor, once the new equilibrium is reached, will be distributed between the various industries in such a way that its marginal productivity is equal for all. This result is contrary to the first condition, and thus in the case considered we cannot have a supply curve of a commodity under conditions of increasing costs. The supply of corn is typical of this case. An increase in demand provokes an intensification of cultivation and thus an increase in the cost of corn. *[324]* But to a similar degree the cost of other agricultural products, that are possible substitutes for corn, must increase (even if the quantity of them produced remains the same), and this leads to a fresh modification of the demand conditions for corn, which were based on the possibility of obtaining substitutes at a lesser price.[1]

In the contrary case, in which the number of industries using a common factor is very large, we could not accept that the increase in production of one of these has as its effect an increase in the cost of all industries without supposing that the variation in the quantity produced by it should be considerable, which would be contrary to the second condition. A small increase in the production of a commodity would have negligible effects, both on the cost of the commodity itself, and on the cost of the other commodities of the group. The supply of the product must, therefore, be considered as being conditions of constant costs.

The substance of the argument rests on the fact that the increase in production of a commodity leads to an increase in the cost both of the commodity itself, and of the other commodities of the group. The variations belong to the same order of magnitude, and therefore are to be regarded as being of equal importance. Either we take into account these variations for

[1]The difficulty arises in the case of the supply curve of corn (in the literal sense of the word), that is, of one among the different products of the land. It does not entirely invalidate Ricardo's law of decreasing productivity of the land, even if he expresses it in terms of corn: 'the term *corn* was used by them [English classical economists] as short for agricultural produce in general, somewhat as Petty (*Taxes and Contributions*, ch. XIV) speaks of "the husbandry of Corn, which we will suppose to contain all necessaries of life as in the Lord's Prayer we suppose the word Bread doth".' (Marshall, *Principles*, p. 509, note 2).

all industries of the group, and we must pass from the consideration of the particular equilibrium of a commodity to that of general equilibrium; or else those variations in all industries are ignored, and the commodity must be considered as produced under constant costs. What is inadmissible is that the equal effects of a single cause are at the same time considered to be negligible in one case, and of fundamental importance in the other. However, it is necessary to accept this absurdity if one wishes to give a general, and not an anomalous character, to the supply curve of a product under conditions of increasing costs.

The inadmissibility of the supply curve of a product (under conditions of increasing costs) in *[325]* the manufacture of which factors are required that are also needed in the production of other commodities, has been maintained by Barone.[1] But he used a different argument from the one we have followed. His argument has been subjected to criticisms that seem to us justified. Since we substantially accept Barone's conclusion it seems to be necessary to show how these criticisms are not applicable to our argument. Barone maintains that under those conditions the supply curve of a product cannot be formed, because its cost is a function, not only of the quantity of the product itself, but because of the quantity of other commodities produced in which the fixed factor appears as an input: 'It is certainly true that for each product, by making the hypothesis that the quantities of all the other products remain unchanged at their equilibrium level, a supply curve can be constructed',[2] but such a curve would be useless in the determination of a particular equilibrium, even an approximate one, because that hypothesis is not close enough to reality. As Ricci pointed out 'the observation is correct but proves too much, because the demand for a commodity (A) is also a function of the price of (A) and, jointly, a function of the price of the other commodities (B), (C), . . . , and thus, strictly speaking, demand curves for commodities should not be considered as functions of one variable',[3] as are used by Barone. The objection hits the mark, for the hypothesis rejected by Barone in the case of supply, is effectively of the same degree of approximation as the one accepted in the case of demand. He himself, foreseeing the criticism, had tried to defend himself by saying that 'having

---

[1]'Sul trattamento di quistioni dinamiche' *Giornale degli economisti*, 1894, II, p. 425, *et seq.* Later, Barone changed his mind, and included the supply curve of a product in his theory; but, like Ricci, he confused it with the particular expenses curve (see above pp. 19, 39, note 1, and 40, note 1).

[2]p. 427.

[3]*Loc. cit.*, p. 224.

made a first hypothesis, so as to simplify the problem, is not sufficient reason for making another and in this way renouncing all the approximation that one can reach on the basis of the assumption already made'.[1] This is imprecise, for the second assumption is no less approximate than the first, and therefore there is nothing to be gained by giving it up. *[326]*

But our argument is not concerned with the greater or lesser approximation of the assumption that the prices and quantities of the other commodities which use a factor in common with the commodity under consideration, remain unchanged. Our argument is that that assumption is absurd, and contradicts the preceding hypothesis, for the increase in production of a commodity leads to an increase in cost that has equal importance for that commodity and for the others of the group; so that it cannot be taken into consideration for one and ignored for the others. This argument, which leads us to the conclusion that cost is to be regarded as constant, is perfectly compatible with the hypothesis accepted for forming a demand curve for a commodity, that is, that the marginal utility of money for a consumer does not change with the variation in the sum spent by him on one among the many commodities he buys, and that therefore the quantity and the price of the others do not change. We are here dealing with quantities of a different order of magnitude (the variation of the marginal utility of the commodity considered and the variation of the marginal utility of money in relation to a variation of the former) and the quantities of the second order of magnitude can be ignored, whilst taking into account those of the first order.[2]

The consequences of these conditions are just as serious for the supply curve in circumstances of increasing costs, because these conditions also imply that variations, which are of equal magnitude and originate from an identical cause, are considered to be negligible on the one hand and of importance on the other.

It is necessary that the advantages of increased production in the industry considered should not have repercussions in any way on the other industries. The economies of large scale production must be 'external' from the point of

---

[1] *Loc. cit.*, p. 329.

[2] *Cf.* Marshall, *Principles*, p. 132; and Barone, *Giornale degli economisti*, 1894, Vol. II, pp. 217, 221, 416. The decreasing utility of money (which is the 'factor of production' of all utilities) and the decreasing productivity of land (which is a common factor for all agricultural products) must be ignored for the same reason when we consider demand and supply of a single commodity.

view of the individual firms, but 'internal' from the point of view of the industry. It is a question of seeing within what limits it is reasonable to suppose, on the one hand, a close interdependence among firms in an industry, and, on the other hand, an absolute independence of the same firms from producers of other commodities. If we investigate what these external economies really consist of, *[327]* we find that very few of them possess such a qualification.[1] The most important ones, if indeed they do derive in part from the development of a single industry, are generally to the advantage of all the industries found in the district in which the development is taking place. This is especially true for those basic external economies 'which result from general progress of industrial environment',[2] and for those deriving from the development of means of communication and transport.[3] Marshall, who in his *Principles* has given such great weight to external economies peculiar to a single industry recognised, in the book in which he wished to approach nearer to reality, that 'the [external] economies of production on a large scale can seldom be allocated exactly to any one industry: they are in great measure attached to groups, often large groups, of correlated industries',[4] a correlation that can both consist of territorial proximity, and of an affinity of products. External economies of this sort cannot determine a tendency towards decreasing costs that would satisfy the required conditions. With the increase in production of a commodity, if it utilizes a large part of the resources of a country, the prices of very many other commodities will decrease, and thus the static system which is a necessary premise of the supply curve is overturned.

When this difficulty does not present itself, another difficulty remains which hinders the application of external economies common to different industries to the supply curve under conditions of decreasing costs. This difficulty arises when the industry considered appropriates only a small part of the resources of a country, and this, in order to exert an appreciable influence on the totality of the other industries, it must undergo a great change. But the

---

[1]Among the external economies that possess this qualification, the only really important ones are the formation of a market for the sort of work particularly sought after by the industry considered, and the best organisation of the market for its products. But these things cannot be taken into account in a theory that contains among its premises perfect competition, that is, which presupposes, right from the start, a perfect organisation of the markets.

[2]Marshall, *Principles*, p. 441.

[3]*Ibid.*, p. 317.

[4]*Industry and Trade*, London, 1919, p. 188.

supply curve is of significance only for small variations in the output of an industry. Thus we cannot, *[328]* without betraying the guiding principle of the method being followed in this type of analysis, suppose that the supply curve has a negative inclination only as a consequence of a form of external economies on which small variations in an industry have negligible effect. For example, it is going too far to suppose that a small increase in production of *one* among many commodities can have as a result such an improvement in means of transportation that in its turn reacts in such a way as to make the price of that same commodity decrease. Yet, if that did happen, the prices of all other commodities would decrease at the same time. The argument based on external economies has been little studied from the point of view of concrete reality, and it would therefore be difficult to make a criticism of it from this point of view. But it seems probable that there must be very few cases indeed of external economies which can be introduced as a consequence of a variation—not a very large one—in the size of an industry.

There are then strong reasons, of which we have tried to show the most prominent, why, apart from exceptional cases, non- proportional cost curves cannot be involved in the determination of the particular equilibria of single commodities in a static system of free competition, without assumptions being introduced that contradict the nature of the system. An essential condition is to totally isolate the industry that produces the commodities considered from all other industries. Now, for increasing costs, it is necessary to take into consideration the whole group of industries that uses a given factor of production. For decreasing costs we must consider all groups of industries that reap an advantage from certain 'external economies'. These causes of variation of cost, highly important from the point of view of general economic equilibrium, must of necessity be considered to be negligible in the study of the particular equilibrium of an industry. From this point of view, which constitutes only a preliminary approximation to reality, we must then concede that, in general, commodities are produced under conditions of constant costs.

PIERO SRAFFA