The Great Debates on the Laws of Returns and the Value of Capital: When Will Economists Finally Accept their own Logic?

Despite long years of important studies and critical debates, economists who have come to grips (or are beginning to do so) with the two issues in the title of this article are few and far between. The criticisms have been ignored or relegated to the category of “paradoxes”, so as to put them on one side and thus justify the failure to give up outdated approaches.

A thorough examination of this state of affairs is called for.

1. Ricardo on machines

According to Ricardo, wage increases stimulate the substitution of machines for workers since, when wages go up, there is no change in the price of machines. Wage increases involve a cut in profits, since wages increase for all the capitalists and not only for those producing machines. The latter group therefore, cannot put up their prices in order to prevent the reduction in profits. (Ricardo develops this argument in section V of Chapter I of the Principles.) Subsequently, after the labour-saving machines have been introduced, which involve an increase in productivity, there will, therefore, gradually be a general fall in prices, which will lead to a rise in the purchasing power of all

---

8 I express my thanks to Jan Kregel, Marcello Massari, Luigi Pasinetti, Sergio Ricossa and Alessandro Roncaglia for their comments and critical observations. My thanks do not necessarily imply agreement of these economists with the views here put forth; this, Sergio Ricossa certainly does not agree with the main propositions of my paper, although the differences are less deep as they might appear at first sight. The English translation of this article has been financed by a grant of the Consiglio Nazionale delle Ricerche (no. 86.01252.10).
economic operators, though in differing proportions. On the basis of this reasoning, Ricardo concludes that, with the increase in wages, the ratio of wages to the price of machines tends to rise, and, more generally, that, the shorter the durability of fixed capital, the greater the increase in the relative prices of the goods. In Ricardo’s reasoning, the substitution of machines for labour becomes expedient either when wages increase or when, while wages remain unchanged, a new machine is invented which is more efficient but not more expensive than the machines already available. The second case is equivalent to the one in which there is a fall in the price of the machines already available. In both the former and the latter case, there is an increase in the ratio \( W/P_m \). Ricardo examines the former case in the first chapter of his Principles (pp. 40-43). The second case is studied in his famous Chapter XXXI on machines, in which he postulates the possibility of technological unemployment. It would appear that Ricardo regards the two cases as being equivalent. For, towards the end of Chapter XXXI, he observes: “machines and labour are in constant competition with each other, and often the former cannot be used until labour becomes dearer”. (“Often”, therefore, but not necessarily; machines, if they are new and more efficient than those already on the market, can be introduced even if wages do not rise.) Ricardo makes it clear that, in examining the introduction of machines, he is referring to an irreversible process which takes place over time gradually and which involves changes in technology. Hence, the substitution between machines and labour which he discusses is a dynamic process which, despite certain appearances, has practically nothing to do with the static substitution of marginal analysis. Admittedly, anyone used to reasoning in terms of that analysis will find it hard to realize this, and will be tempted to interpret the very concepts of substitution and productivity in those terms. But we must not be misled by appearances. The picture becomes clearer if we think over the two concepts just mentioned, beginning with the second one.

2. Average and marginal productivity

The average productivity of labour is given by the ratio of total production and number of persons employed: this is a concept which does not give rise to problems either in static or dynamic analysis. Problems emerge with the concept of marginal productivity: assuming that production can be represented as a function of a certain number of factors, marginal productivity is the partial derivative of production to be related to a very small change in one factor, on the assumption that the others remain constant. Mathematically speaking, the concept is undoubtedly valid. But is that also the case from the economic point of view?

The reply is that it is not. The fact is that any technology is embodied in specific types of fixed capital which make possible specific combinations of factors and not others. If, in a textile factory with a given plant, it is desired to produce a square inch more of cloth, we need not only to increase, by however small an amount, the degree of utilization of the plant (assuming that this is possible), but also to have at least one worker a little longer, who, unless he is prepared to weave the fabric out of air, will have to use a wad of raw cotton — if we are talking about a cotton cloth.

Marginal productivity conceived of as a partial derivative therefore turns out to be a concept devoid of any real relation with economic reality. Not for nothing did Marshall, who had the knack of minimizing the difficulties inherent in certain theoretical models in order to make them fit in with reality, and never more than when he tried to justify the marginalist theory of income distribution, refer, not to marginal productivity in the strict sense of the term, but to the net product of a factor, defined as the additional output obtained by an incremental application of that factor “after allowing for incidental expenses” (Marshall, p. 432). Now, the “net product of a factor” and the “marginal product” are not at all the same thing. A partial derivative is a partial derivative, full stop. To persuade himself that he was right, even before persuading the reader, Marshall furnishes one of the rare examples able to give a veneer of plausibility to the equation of “marginal productivity” with “product net of incidental expenses”: the example is that of a sheep farmer who wonders whether it is worth while taking on an additional shepherd — a decision which can disregard the possibility of “any additional expenditure on plant or other means of production” (p. 427). In fact, in order to illustrate the concept of a factor’s marginal productivity — usually labour — it is the almost invariable custom to use examples of this type, taken from agriculture or sheep-farming, since these are examples in which the marginal productivity of one factor — the amounts of the other ones being equal — may
appear not manifestly absurd, given the modest amount of the means of production which the additional worker has to use to contribute to productive activity. But what is being discussed is a partial derivative, so that, even in cases of this kind, it is not permissible to forget that the additional shepherd boy — to go back to Marshall's example — needs not only a bed and food, but also some rudimental equipment to help him tend the sheep. The insignificant items, which may be disregarded in other contexts, are not to be ignored in this case. If that is true in the special example in question, it is much more true when we are considering the normal circumstances of modern industry. When it is recognized that the receipt imputable to the increase in output due to the new worker must be considered net of the charges relating to the other means of production, it will be admitted that, for every technique, the factors are necessarily complementary to each other, so that it appears that we rule out the possibility of involving the concept of a partial derivative, and the whole construct, which is based on that concept, collapses. It also collapses for a more particular reason. Marginal and average productivity diverge when, in a diagram representing the behaviour of production in relation to that of a variable factor, the former does not take the form of a straight line. If it is a straight line starting from the origin, then marginal and average productivity coincide. Now it appears that this is precisely the normal case. In other words, given the technology, constant returns seem to be the rule. Diminishing returns appear to be an artificial construct, worked out by economists in line with the hypothesis of generalized competition, a hypothesis regarded as vital by traditional economic theory. Ricardo had considered diminishing returns referring to a long period and to a country where the population was increasing and not to an individual product but to a whole branch of economic activity (agriculture) using a limited factor (land). On these conditions, but only on them, the potential trend towards diminishing return may be accepted, even if in real life, technical progress, contrary to Ricardo's belief, has more than offset this tendency, so that in agriculture, too, increasing returns are the rule. Now the extension to the individual firm

or to the single factor of the principle of diminishing returns, on the basis of the idea that such trend obtains whenever a factor varies while at least one of the other factors remains constant, is an extension which is inadmissible. In the short run, the constant factor for the firm would be given by plant and machines. But every machine, when used, calls for a fixed number of workers. Moreover, the degree of utilization of a plant may vary, but that affects the level of total average cost, and not the returns of the variable factors. As a rule, the productivity per hour worked remains constant when the degree of utilization varies. Not even in the long term (the plant is no longer given) does the principle of diminishing returns hold good. I would refer readers to the monograph, Oligopoly and Technical Progress (pp. 26-9, 39a and 162) and to my Lezioni di economia (Vol. II, part. IV, Chapter II, first section).

Nobody denies that the means of production are complementary to each other. But it is believed that, within this general complementarity, each technique allows for margins of substitutability. In particular, it is asserted that, whereas in chemistry the law of definite proportions holds good, this does not hold for economics, since one and the same product can usually be obtained by different combinations of factors corresponding with different techniques. This is true, but it is also true that each technique involves definite proportions, that is, imposes a rigid complementarity between the factors: the possibility of different combinations arises only when we are considering different techniques; but this possibility does not involve that of isolating the marginal contribution of a single factor. To clear up this point once and for all, let us examine very slight changes in the combination between "capital" (whatever the meaning of the term) and labour — K and L — and production, X: let us examine three hypothetical cases: the first is that of the marginal productivity of one of the two factors (let us say, labour), the second one illustrates the complementarity nexus, and the third the substitutability nexus. Well, the first case is not admissible, but

---

1 When Ricardo (1981, p. 14) writes: "the marginal productivity of labour may take the form of output per worker minus capital inputs per worker needed to create the job for him," he is putting himself on the same plane as Marshall, and in this way is giving up the concept of the partial derivative which presupposes the rigid constancy of the other factors: a precondition, it should be noted, which is recognized (in contrast to this quotation) in a later passage (p. 44, half way down).
the second and third ones are (the third case of necessity implies a change of techniques):

\[
\begin{array}{ccc}
(I) & (II) & (III) \\
\Delta X & 0 & + & + \\
\Delta L & + & + & - \\
\Delta X & + & + & + \\
\end{array}
\]

With reference to the third case and to repeat the earlier observations, let us say that, while, in the production of chemical compounds, there are definite and unique proportions, in the production of goods we find definite but alternative proportions in relation to different techniques.

If it is not possible to speak of a factor marginal productivity since each technique is conditioned by a strict complementarity, we must conclude that it is not permissible to derive the demand curves of the individual factors from their marginal productivity curves, since these demand curves simply do not exist. What does exist, in relation to the factors of production, is the problem of joint demand. In his Principles, Marshall devoted a no means fleeting attention to the problem, and had suggested that the demand curves of the factors should be constructed following the criterion of joint demand, which is a very different matter from the criterion of the “net” marginal product. The two approaches are radically heterogeneous, since the former is based on the complementarity of factors, and the second on their substitutability (which is admissible, it should be repeated, only with reference to different techniques). The heterogeneity of the two approaches does not at once appear evident since the two analyses are sharply separated: Chapter VI of Book V, “Joint and composite demand”; Chapter I of Book VI, “Preliminary survey of distribution”; — the reference is to the distribution of national income which, as it happens, forms the title of Book VI. 3,4

3 As regards the demand curves obtainable from the “net” marginal product, I have already expressed a few comments; as regards the demand curves obtainable from the joint demand, the following quotations are worth thinking over: “The demand for raw materials and other means of production is indirect and is derived from the direct demand for those serviceable products which they help to produce” (p. 316) “... The demand schedule for any factor employed in the production of a given commodity can be derived from that for the commodity by subtracting from the demand price of each separate amount of commodity the sum of the supply prices for corresponding amounts of the other factors” (p. 318).

4 On the question of the complementarity and the substitutability of the factors, see L. Pasinetti (1984, pp. 201-4).

3. Labour-saving inventions

Thus, the concept of marginal productivity raises insuperable difficulties, whereas the concept of average productivity raises none. It can be expressed either by a ratio of overall output to the amounts of a specific factor, or by the inverse ratio, which is nothing but the coefficient of production. In general, the average productivity of each factor (having discarded the concept of marginal productivity, we can talk simply of “productivity”), if we take technology as given, appears to be constant; while, if we admit that techniques vary over time, it is generally increasing. If a certain factor becomes dearer, there is an advantage in introducing new techniques which increase its productivity, that is, which economize it in terms of the unit of output. Given the complementarity between the different factors, the entrepreneur, if he is not successful in economizing in the use of the factor which has gone up in price, may seek to economize in the use of other factors. However, it is characteristic of modern capitalism that labour is the factor which tends systematically to become dearer and dearer. For this and other reasons, labour-saving inventions, that is, inventions which reduce the labour input, are decidedly more frequent than those which save fixed capital, energy, raw materials and intermediate products, because, especially when these factors become dearer, there is no lack of inventions which permit savings in their use.

With reference to a specific good, an invention can therefore be defined as labour-saving simply if there is a reduction in the input of direct labour. However, the invention is of economic importance and is only adopted if, in the type of production considered, this reduction leads to a greater saving in wages than any additional charge due to the machine and other means of production. 2 In general, an increase in the ratio of wages to the cost of the machines may stimulate the flow of labour-saving inventions; this may happen either because wages rise or because the cost of the machines falls (as a result of specific innovations), or because wages rise more than the cost of the machines. If wages increase, this increase affects all types of production, including

5 Pasinetti defines technical progress by measuring all the inputs in terms of direct and indirect labour. For the purposes of the present essay, there is no need to consider Pasinetti’s concept, which is broader than the one suggested in my text, and which is analytically complex, since indirect labour raises the question of time, and hence of the simple and compound rate of profit to be applied to the different quantities of labour (L. Pasinetti 1984, p. 208).
4. Dynamic and static substitution

Static substitution in traditional theory involves, in equilibrium, the equality of the marginal rate of substitution (which is negative) and the ratio of factor prices, say labour and "capital":

\[ \frac{dK}{dL} = \frac{W}{P_k} \]

Hence, if wages increase relatively to the price of "capital", a given type of production is carried out in which the proportion of "capital" to labour becomes higher than in the initial situation — where initial situation, change in wages and final situation are expressions excluding any temporal element: they are hypothetical, instantaneous and reversible changes.

In the case of dynamic substitution, on the contrary, we have an incentive to accentuate the mechanisation of the processes of production when there is an increase in the ratio of wages and the price of machines, whatever the reason for this increase. If we call \( L \) the labour input, \( W \) the wage rate and \( P_m \), the price index for machines, we have the relation

\[ I = a' - b'W/P_m \]

that is, since the inverse of \( I \) indicates average labour productivity

\[ \pi = a'' + b''W/P_m \]

or again, recognising that the substitution of machines for labour goes on almost without interruptions,

\[ \pi = a'' + b'W/P_m \]

where the cap over a variable indicates a rate of change over time (usually a year).

The formal similarity between the concept of static and dynamic substitution remains. There is in fact a nexus between a very slight change in the labour used (dL) and the labour input (L), and the wage rate appears in both concepts. On the other hand, the cost of the machines \( (P_m) \) comes out as similar to that of "capital" \( (P_l) \). However, in the relation regarding the dynamic substitution, does not appear any term implying a slight change in "capital". But the differences go much further. The essential difference lies in the basic concept: the marginal rate of substitution is regarded as the ratio of the marginal productivities of the two factors:

\[ \frac{\partial X}{\partial L} = \frac{dK}{dL} \]

but, in the view put forward here, the idea of marginal productivity is inadmissible. Moreover, "capital", which in traditional theory has an ambiguous significance and which as a rule also includes working capital, is represented here by the machines, and the cost of capital here is not given by the rate of interest, but by an index which expresses the price of the machines. On the other hand, in traditional theory the changes are theoretically reversible, while here changes take place over time and hence, among other things, are irreversible. Lastly, in traditional theory the different techniques are assumed as being already known and available, while here it is the progress of the techniques which is being analysed, a progress set in motion by the ratio of wages to the price of the machines and by other impulses.

If the concept of marginal productivity is not acceptable, that of the marginal rate of substitution \( s \). It can be imagined that a given level of production can be obtained by different combinations of "capital" and labour, so that, where the former rises, the second falls. In fact, it is
not unreasonable to assume that, switching from one combination to another, the other factors — energy, materials and intermediate products — remain unchanged. (On the hypothesis based on marginal productivity, we must on the contrary assume that all the factors except the one whose marginal productivity it is desired to measure, and it must be admitted that, as the result of a very slight increase in that factor, production may increase, even if only infinitesimally, and that is not admissible.) The mathematical expression itself shows that the concept of marginal rate of substitution is, per se, acceptable, while that of marginal productivity is not. There is a simple derivative in the first case, and a partial derivative in the second one.

Since the well known classification by Hicks of inventions into labour-saving, neutral and capital-saving categories is based essentially on the concept of marginal productivity, this classification cannot be accepted. As we noted, an invention may be defined as labour-saving — or as economizing other factors — simply when it reduces the labour input or some other input. From this point of view, it is meaningless to speak of neutral inventions. If anything, we can argue about inventions which do not permit of any factor saving, but give rise to new goods which were not previously produced.

5. Pasinetti and dynamic substitution

As we have seen (cf. the third column of the table in section 2), static substitution is admissible on the logical level. In fact, however, it is of only slight significance. The techniques which can be used to produce a specific quantity of a certain good with different combinations of labour and physical capital are few in number. In certain cases, there are alternative techniques only at different levels of production. To be more precise, there is presumably a wide range of techniques when we consider both the basic processes and the subsidiary operations. But the fundamental processes and the various fixed capitals which usually embody them are few in number. Therefore, the scope for static substitution appears to be circumscribed. The fact is that the techniques to be considered must be the ones we are familiar with: the introduction of new techniques in traditional theory, which is strictly static, is excluded. But the substitution observed in the real world is almost always linked to technical progress, either because the ratio of wages to the cost of machines, as we have noted, stimulates labour-saving innovations, or because innovations of this type are introduced independently of direct economic stimuli. The increase in the ratio $W/P_m$ constitutes only one of these stimuli. This ratio may increase because, the prices of the machines remaining constant, wages increase as a result, for example, of action by the trade unions. But this ratio can increase, on the contrary, because, wages remaining constant, the price of the machines falls because of innovations in their production. On the other hand, the ratio $W/P_m$ may increase without necessarily pushing down the profits in the sector producing the machines, because productivity rises systematically. In other words, the increase in productivity appears to be the effect, but also the precondition, of a sustained increase in the ratio $W/P_m$ — and, on a broader scale, in the ratio $W/P$, where $P$ is the general price level.

I have already noted that the similarities between static substitution in traditional theory and dynamic substitution are purely formal. In fact, in the case of static substitution, an increase in wages involves, production being equal, on the one hand a fall in labour intensity ($L/X$), and, on the other, an increase in capital intensity ($K/X$); in addition, an increase in wages involves a reduction in the rate of profit (which, in equilibrium, is equal to the rate of interest), and it is precisely this which leads, according to this theory, to an increase in capital intensity. But Pasinetti is perfectly right when he notes that:

"...there is nothing of the sort that emerges as the usual pattern of technical change. ‘Roundaboutness’, if we take it in the sense of degree of capital intensity, need not change at all, and the rate of profit need not change either and yet the degree of mechanisation (the ‘substitution of physical capital for labour’) may be increasing all the time".*

Pasinetti concludes that, among other things, “the wage rate will increase but as the effect, not as a cause, of technical change”. This conclusion seems to diverge from mine, since I maintain that technical progress is stimulated by, among other things, wage increases. But I am considering the increase, not of wages as such, but of the ratio of wages to the price of the machines; and I am asserting that the increase in that ratio constitutes one of the various economic stimuli, not of technical progress in general, but specifically of labour-saving innovations. In any

---

case, I have already observed that the whole process must be viewed as the result of an unceasing series of interactions between an increase in that ratio and an increase in labour productivity. This does not at all mean that this process concerns all innovations, and that all innovations are implemented as a consequence of direct economic stimuli. In other words, there are innovations — and important ones — which may be regarded as exogenous to the system of production; and it is to these innovations that Pasinetti's assertion cited above fully applies. As regards the other innovations — those stimulated by the increase in the ratio of wages to the price of machines, by the expansion of demand and other direct economic stimuli — that statement must be related, to my way of thinking, to the interaction process mentioned earlier, in which the increase in that ratio is not only an effect but also a cause of the increase in productivity. The divergence — if there is one — is therefore a minor one. The analysis carried out by Pasinetti in the work in question, and I am not referring only to dynamic substitution (Pasinetti prefers to speak of "change of techniques"), is one with which I am in complete agreement.

6. The myth of increasing marginal costs

The unceasing technical progress is the source of the increasing returns which characterize modern productive activity. These are therefore dynamic increasing returns. Only economies of scale due to technologies already known may be regarded as at the origin of static increasing returns. It is not easy to distinguish between the two types of increasing returns. However, it may be confidently asserted that the former are by far the most frequent and the most important in modern economies. Static increasing returns only take place occasionally, and not infrequently through processes of imitation between different countries.

If we recognize that increasing returns of the dynamic type are predominant in modern economics this does not mean that we deny that in the short period, and given techniques, returns are as a rule constant — constant in relation to changes in production. In a diagram we indicate on the x-axis the amount produced (X) and on the y-axis the average labour productivity (X/L); this, if technique is given, may be shown by a straight line parallel to the x-axis. Dynamic increasing returns may be shown by a series of straight lines, each parallel to the others, which gradually shift upwards. Each straight line is dated and corresponds to a specific technology.

\[
\begin{align*}
\pi_1 \quad & \pi_2 \quad & \pi_3
\end{align*}
\]

If the (average) productivity of the variable factors is constant, the marginal cost coincides with the direct or variable costs, and the average total short-term cost declines up till the limit of the plant's capacity (the average total long-term cost falls, if there are economies of scale, otherwise, it too is constant).

The essential aspects of this state of affairs were accurately analysed by Sraffa as long as sixty-two years ago (1926, p. 543):

"Everyday experience shows that a very large number of undertakings and the majority of those which produce manufactured consumers' goods work under conditions of individual diminishing costs. (...) Business men, who regard themselves as being subject to competitive conditions, would consider absurd the assertion that the limit to their production is to be found in the internal conditions of production in their firms, which do not permit the production of a greater quantity without an increase in costs".

Despite Sraffa's scathing criticism ("...would consider absurd the assertion...") most economists have turned a deaf ear to his arguments. Vested theoretical interests carried too much weight. Sraffa was followed by Kalecki (April 1938). But economists have, undaunted, continued to teach their U-curves. The picture drawn by Sraffa is entirely consistent with the empirical indications set out by Dunlop (September 1938) to show that Keynes' conviction expressed in the General Theory
(that, because of diminishing returns, when money wages increase, real wages tend to fall) is unfounded. This conviction, as Keynes observed in his rejoinder to Dunlop "was largely shared by English economists" (and not only by them). We have then the results of the survey carried out by Hall and Hitch (May 1939), which clearly presupposes constant direct costs, and hence their coincidence with marginal costs. From then on, the bulk of empirical indications disproving diminishing returns and favouring constant returns (techniques being equal), or increasing returns (because of technical progress) has been huge. Despite all this, economists — with few exceptions — have not batted an eyelid. Some of these indications have been conceptualized, and others not. But there is no question as to the conclusions to be drawn from them. Here, in a telegraphic style, is a list of the indications to which I am referring:

1) Verdoorn’s (or “Smith-Verdoorn’s”) Law;
2) Okun’s Law;
3) Generally satisfactory econometric results of the relations between direct costs and prices (if direct costs followed a U path, there would be no systematic relation between direct costs and prices).
4) The almost general use, and not only in empirical analyses, of the concept of the degree of utilized capacity, a concept which would be meaningless if direct costs followed a U path.
5) The generalized use in the cost accountancy of industrial firms of criteria which presuppose constant direct costs (such is the implicit assumption in the diagrams aiming at identifying the break-even point — the break-even charts).
6) The generalized adoption by the firms acting as price leaders of criteria based on a mark-up over direct costs: such criteria, too, presuppose constant direct costs.

Empirical analyses for the most different kinds of industry show that, in the vast majority of cases, short and long-term marginal costs are constant or decreasing. Nevertheless, Mansfield, the economist who has taken the trouble to assemble the results of these analyses, concludes, under the influence of traditional theory, that “it is inconceivable that in the end marginal costs will not increase with the quantity produced” (Mansfield, 1975, pp. 189-196). A fine act of faith — in flat contradiction with the well-known Anglo-Saxon pragmatism! It should be pointed out that the device with which authors seek to find a rational justification for their act of faith is usually that of supposing that marginal costs are approximately constant over a considerable interval, which is the one which can be empirically observed: as we near the full capacity of a plant, the curve of these costs would tend to soar. Here are the two curves: the one invented by economists in homage to the needs of the traditional theory — or else to avoid polemics regarded as useless — and the one which has to be constructed on the basis of empirical research:

The difference would lie in the initial section of the steep increase AB, which would occur in case I, but not in case II; but, since the section AB remains invisible, we have the right to ignore it. This is on the empirical plane. On the theoretical level, we have the duty to ignore it, since the arguments seeking to prove the increase in marginal costs do not stand up to criticism.

So far, economists have remained unmoved by the impressive and increasing number of theoretical criticisms and empirical analyses which I have just referred to. The moment has come, I feel, to slough off this seraphic and untenable imperturbability.

---

7 Cf. F. Mombriamo, 1975, pp. 127-8. (Mombriamo is one of the few economists who combines a thorough theoretical mastery of the subject with a considerable body of direct experience in industry.)
7. The cost of capital and the rate of interest

In my opinion, the moment has also come when economists should abandon their imperturability as regards another fundamental question; that of the cost of "capital", a question connected with the previous one via the concept of "marginal productivity".

If a theory is wrong, criticisms of it may be different, and, at the same time, all sound. When a boat, as we say, is leaking everywhere, it is important to find the holes one by one only because the sailors are insisting in denying the existence of any holes. On this fundamental question, too, Sraffa has provided an important critical contribution. The specific criticism, which should be seen in the context of his model of general analysis, is summed up in the proposition that capital cannot be regarded as a quantity which can be measured independently of income and price distribution.

Where there is confusion as regards ideas, there is bound to be ambiguity, too: this has always been the hallmark of the concepts of "capital" and the "price of capital". "Capital" is sometimes taken to mean fixed capital and sometimes both fixed and working capital, but in most cases it seems to have the former meaning. Faced with the logical difficulties arising precisely when they came to consider the question of capital as a given quantity, some economists thought they could settle matters by a witticism — capital conceived of as a jelly; but witticisms cannot take the place of theoretical analysis.

The ambiguity reaches its height when economists in the neoclassical tradition examine the problem of income distribution. Here, there emerges the concept of the aggregate production function which involves very right constraints on interpretation. There is room only for a quantity of labour \( L \), viewed as a mass of workers who can be rendered homogeneous, and for a capital \( K \), conceived of as a quantity given \textit{a priori}, a capital which would seem to include both fixed and working capital and the price of which is stated to be given by the rate of interest, which in equilibrium is equal to the rate of profit. It should be noted that, in this context, the consideration of any "incidental expenses" would rule out the possibility of using the aggregate production function to explain distributive shares.

Now, we should give thought to the point of view of business men who consider capital in financial terms, that is, as purchasing power, and interest as the price they must pay to obtain it if internal funds, accruing from profits, are insufficient (when they use internal funds, they must also in their accounting enter the interest paid as one charge among others). It follows that, when the interest falls, there is no capital substitution (in physical terms) for labour. On the contrary, there is an incentive to purchase additional means of production in order to expand output, since, with cheaper financial funds, one can purchase not only the material means of production but also the services of the workers. It follows, too, that it is no wonder if a gradual fall in the rate of interest is accompanied by repeated switches in techniques, both towards the increase in the quantity of capital per worker and in the opposite direction. This is the problem of the "switch of techniques" which was first formulated in Sraffa's book, \textit{Production of Commodities by Means of Commodities} (Chapter VI). It should be noted that, in the first model set out in this work, labour does not appear as such. What do appear are the means of subsistence which are equivalent to wages and which cannot but be advanced; in this case — which is the only one considered by the classics — it is at once obvious that the rate of interest affects both the material means of production and the services of the workers. But, even if we adopt the second model, in which wages are paid \textit{post factum} and form part of the surplus together with profit, the conclusion is analogous; only the road to be followed is longer. And the conclusion is that, in the problem of the choice of techniques, there is no point in considering wages as being in opposition to interest. (Since the classics argued exclusively in terms of Sraffa's first model, in which wages are advanced, they would not even have been able to understand the problem of the choice of techniques as it is debated in our day.) If we then wished to extend Sraffa's second model, by moving (to use Marx's terminology) from simple reproduction to expanded reproduction, the hypothesis that wages are advanced would again become relevant, and the criticism of the traditional theory would again be immediate; since we would have to regard as normal the case of entrepreneurs who, being unable to finance the expansion with internal funds, also have recourse to borrowed funds in order to take on additional workers, as well as additional material means of production.

If it is not interest, what should we consider as "the cost of capital"? But, naturally, the price, or rather the prices, of physically determined capital goods, and, in particular, of fixed capital. In this way, however, we cannot use the aggregate production function to explain income distribution.
8. Income distribution in the marginalist theory

If it is true that the factors of production are complementary to each other, but, for each of the possible techniques, not substitutable, then it is not possible to distinguish between average and marginal productivity. If, the technique being given, average productivity is constant, which seems to be the rule, then, from the mathematical point of view, average productivity and the hypothetical marginal productivity coincide. And it is no longer possible, even formally to derive from marginal productivity, which is equal in each case to average productivity, distributive shares. If we relate the amount produced, \( X \), a given factor, let us say labour, \( L \), and if the relation can be represented by a straight line starting at the origin, the derivative \( \frac{\partial X}{\partial L} \) is equal to the ratio \( X/L \). From the aggregate production function \( X = f(L,K) \), we obtain the relation:

\[
\frac{\partial X}{\partial L} \cdot L + \frac{\partial X}{\partial K} \cdot K = X. \tag{A}
\]

But if \( \frac{\partial X}{\partial L} = \frac{X}{L} \) and \( \frac{\partial X}{\partial K} = \frac{X}{K} \) we would have

\[
\frac{X}{L} \cdot L + \frac{X}{K} \cdot K = X
\]

that is, \( X + X = X \), which is absurd.* The absurdity disappears if it is recognized that the average productivity of a factor also implies the contribution of the other factor or factors: the average productivity of a factor merely registers the changes in production in relation to those in that factor. Thus, when we consider the average productivity of labour, we also implicitly take account of the contribution of capital, and a similar observation holds good for the average productivity of "capital".

---

* Perhaps this is the key to the problem of why, in the empirical estimates of aggregate production functions of type \( Y = A L^\alpha K^\beta \), the sum \( \alpha + \beta \) tends to be equal to unity; this result would express the constant implicit in equation (A).

Introducing the incidental expenses of Marshall (and Ricardo), which, according to the example given in Section 2, would concern the marginal productivity of labour, equation (A) would lose any meaning as regards distribution shares, since \( \frac{\partial X}{\partial L} - \text{E} \cdot L \) would become \( \frac{\partial X}{\partial L} - \text{E} \cdot L \). Strictly speaking, we would have to make an analogous correction in the second term \( \frac{\partial X}{\partial K} - \text{K} \) as well. The incidental expenses, E, should be regarded as indicating the impossibility of neglecting the complementarity of the factors.

9. Income distribution: Sraffa, Kalecki and the mark-up

For Sraffa, income distribution is not determined by purely endogenous forces. If we suppose that income is distributed between wages and profits (the rate of interest and that of profit tending to be the same), and if we regard wages as being above subsistence level, what emerges is one degree of freedom, so that the distribution of the national income becomes indeterminate. On the logical plane as a quantity given externally, we may consider either wages or the rate of profit. Sraffa (1960) gives the reasons for preferring to take as given the rate of profit rather than wages:

"The rate of profits—be it equal to the price of a commodity—has a significance which is independent of what is going on in the rest of the economy. It is accordingly susceptible of being determined from outside the system of production, in particular by the level of the money rate of interest."

10 Cf. for example, F. M. Fisher, 1969.
In Sraffa’s conception, then, the distribution between wages and profits is indeterminate. It would seem that the distribution is determined by economic policy decisions and by conflict. His conception is compatible with what is in the last analysis a view of distribution in which the shares are fought for. In the abstract, we can imagine a system of distribution which ranges from the extreme in which wages are equal to zero to the opposite extreme in which profits are equal to zero. In his book Sraffa considers given quantities of goods, and does not examine the question of the destination of the surplus. When, however, we decide to go on to examine the growth of the system of production, the surplus then assumes a role which is actually vital, since growth can take place if at least a part of the goods forming the surplus are used to increase the mass of means of production. From this angle, that part of the surplus can no longer be freely disposed of, and, admitting that profits constitute the incentive to invest, the source of self-financing and (through various intermediaries) one of the sources of external financing, changes in distribution working against profits will sooner or later block the process of accumulation—a process which, in the real world, takes place through fluctuations. And, since the halting of the process of accumulation, in reality, does not mean that the economy subsides into a stationary state, but brings about an economic recession, which among other things involves a diminution in the level or in the rate of increase in wages, it seems evident that the fall in profits stops long before the increase in wages has wiped them out. In other words, the area of indeterminateness in distribution appears to be severely circumscribed, without disappearing altogether, when, to get closer to actual facts, we propose to analyse the essential aspects of a cyclical process of development. In an analysis of this kind, it is reasonable to admit that the influences “from outside the system of production” may come at times from monetary and credit policy and at times from trade union action and often from a combination of these two impulses, whose weight will vary in the course of time. To tackle such an analysis, however, we must recognize that the rate of profits may be affected, not only by decisions on monetary policy, but also to those imputable to the price policies adopted by firms, who, though to a limited extent, are in a position to have such a policy. To raise this question means to raise the problem of non-competitive market forces, which, at the present time, in the analyses which seek to approximate to reality, are viewed by many scholars as the most frequent ones in modern economies. In the end, the rate of profits may be taken as an “independent variable”; but we must first, in an earlier stage in the analysis, try to explain why it takes on certain values and not others, and why these values change during the cyclical process of growth.

At this point, we have to consider the theoretical models intended to explain the antecedent of the rate of profits, which is given by the mark-up on direct costs. There is now a large number of economists who, explicitly or implicitly, accept the thesis that prices—and in particular those of industrial products—are determined in accordance with some mark-up criterion. Despite this, the theoretical models which seek to explain the mark-up are few in number and rarely debated. Several years ago, I myself put forward a model of this kind, but only a very limited number of economists (Modigliani was one of them) devoted critical attention to it. Yet, once we admit the importance, in the formation and changes of industrial prices, of the mark-up criterion, we can no longer remain on purely empirical ground. Some kind of theoretical model is called for. In an essay dated 1979, I took up this model again and developed it on the dynamic plane (Sylos Labini, 1984, VIII). I then sought to check empirically its implications. Linking up with views systematically put forward by Kalecki (whose works still offer a mine of ideas which have only partly been exploited), I used the proportionate margin to explain the distributive shares in manufacturing industry. Boiled down to its essentials, the logical passages are as follows:

1) In manufacturing industry, the level of mark-up for each product basically depends on the conditions of entry—the entry of other firms and of products from abroad; in their turn, the income conditions depend in the first place on the extension of the market, elasticity of demand, technology, initial advertising expenditure, wages and raw material prices.

2) The aggregate mark-up (expressed as the weighted average of the various margins) does not remain constant in time but tends to vary; as a rule, it varies inversely with changes in direct costs.

3) Starting from the price equation, changes in the mark-up may help to explain changes in distributive shares in manufacturing industry, which forms the dynamic sector par excellence. The changes in the share of other sectors are strongly influenced by the changes taking place in the manufacturing sector. They diverge from them to some extent essentially in relation to: a) the bargaining power of the workers
in the different markets, b) the degree to which the various sectors are exposed to competition (which operates directly only in the case of industrial and agricultural products) and c) whether they belong to the private or the public sector.

The nexus between mark-ups, the rate of profit and the rate of interest come out clearly when the monetary authority raises this rate sharply. In such circumstances, while having to take account of the different constraints (especially those dependent on international economic relations), firms have to raise their mark-ups: those unable to do so may suffer losses, and even go bankrupt.

10. Keynes' marginalism

At first blush, Keynes, a heterodox figure in many respects, seems to appear in an orthodox light in relation to the fundamentals of the traditional marginal theory. I contend that this is not the case, and that we must not be deceived by appearances. I refer in particular, to the two basic questions in the title of the present article. Let us begin with the first one — that regarding the principle of diminishing returns.

In the *General Theory*, Keynes explicitly accepts this principle, and indeed calls for an empirical study of the relation between changes in money wages and those in real terms. He argues that the changes in the two types, by virtue of that principle, would "almost always be in opposite directions". Soon after the publication of that work, two North American economists, Dunlop and Tarshis, carried out research of the kind suggested by Keynes, and came, on the contrary, to the conclusion that changes occur as a rule in the same direction (Dunlop, 1938; and Tarshis, 1939). Keynes (1939) fully realized that the validity of the principle of diminishing returns was at stake. But he was reluctant to give it up, and maintained that, before going to such lengths, more systematic investigations were needed. However, he stressed the fact, that any renunciation of that principle would not have destructive consequences on his analysis and would not weaken, but would even strengthen his expansionist prescriptions.11

11 In an interesting essay, Krugman (1985) brings out the fact that, as early as the 'twenties, Harrod had already realized that, in industry, increasing returns are the rule and that this fact made the expansionist policies by public expenditure more feasible — plans which Keynes had begun to propound well before his *General Theory*. 

Admittedly, the Keynesian hypothesis of constant marginal costs for considerable intervals of the costs curves — as in the path indicated in Graph I in Section 6 — was in any case plausible in the mainly depressed circumstances of the 'thirties. And the price theory, as expressed in Chapter XXI of the *General Theory*, is broadly compatible with the full cost principle, which, on the contrary, is incompatible with the principle of diminishing returns applied to the firm. Many years ago, Bruno Barilli had brought out the fact that "a consistent presentation of Keynes' model is possible only by assuming a type of behaviour of the entrepreneurs based on the full cost principle" (Barilli, 1964). Keynes never took up a clear and unambiguous stand on the question; however, in a rejoinder to Dunlop and Tarshis, he distances himself from the traditional theory on price formation.12 It should be noted that the results of the research by Dunlop and Tarshis, in bringing out the fact that changes in money wages and real wages usually take place in the same direction, do not simply contradict the principle of diminishing returns, but induce one to believe that, as a rule, the opposite principle is valid — that of increasing returns. This "principle", however, cannot be viewed in a static framework, since it is the consequence of technological changes. Adopting a still widely held point of view, Keynes believes that he can assume techniques as given, since the reference is to the short term. But this point of view, which is valid in respect of the abstract short term, is deceptive when — as Keynes does 

---

12 Keynes distances himself with a certain irony and with expressions which echo those used on this point by Staffa. Indeed, it is rare for anyone but an economist to suppose that price is predominantly governed by marginal costs. Most business men are surprised by the suggestion that it is a close calculation of short period marginal cost or of marginal revenue which dominates their price policies. They maintain that such a policy would rapidly lead in bankruptcy anyone who practised it (Keynes, 1939, p. 49).

The businessmen's point of view emerges clearly from the two graphs below. The first of these shows, as tradition demands, two U curves — one of marginal costs, $m$, and one for total average costs, $c$, which include minimum profit. In the second graph, marginal costs are constant and hence coincide with direct or variable costs $p$ is the price, $P$ the temporary equilibrium price.
in discussing the results of the empirical research referred to above — a concrete evolution is being considered, since, while it is true that plants incorporating the techniques cannot be modified rapidly, they may change even in very brief periods because of decisions taken at an earlier date.

Strictly speaking, a price theory compatible with the full cost principle is not compatible with the hypothesis of perfect competition as formulated by traditional theory which calls for a marginal cost which increases from a certain point on. On market forms, Keynes remains deliberately generic: he assumes as given "the intensity of competition", but he does not say whether he considers it as high or low; and, in his reply to Dunlop and Tarshis, he does not at all rule out the view that the most frequent market form in modern circumstances may be imperfect competition (as I have tried to argue elsewhere, the hypothesis of differentiated oligopoly may be regarded as a closer approximation to reality than the hypothesis of imperfect competition). If we indicate the price level as $P$, with $\alpha$ as the mark-up, $W$ as the wage rate and $\pi$ as labour productivity, Keynes' point of view may be expressed by the relation

$$P = \alpha \frac{W}{\pi}$$

This relation holds when unemployment is extensive and the degree of plant utilization is low. When, on the contrary, unemployment dies down, $W$ tends to increase together with the level of employment and productivity tends to fall owing to the action of diminishing returns. (Keynes is considering a closed economy, and therefore disregards the question of imported raw materials, imported finished goods and changes in exchange rates.) Since, in the General Theory, Keynes carries out a large part of his analysis on the hypothesis of widespread unemployment and a low degree of plant utilization, the simple relation indicated above provides, in the Keynesian analysis, the norm. It may be useful to recall that a more complete relation than the Keynesian one can be obtained from a full-fledged theoretical model (the considerations presented by Keynes on prices are essentially fragmentary). The more complete relation follows the same logic as the one followed by the Keynesian relation, without, however, making any reference to diminishing returns.  

---

13 The complete relation is

$$P = \alpha \frac{W}{\pi} + \beta M + \delta P_m$$

where $\pi$ is not constant, $Rm$ is the price index of raw materials, and $P_m$ is the index of imported finished products.

Given all this and while admitting the irredeemable ambiguity of the microeconomic bases of the Keynesian construction, we may recognize that the principle of short-term diminishing returns for industry as a whole is not vital for this construction.

The situation is no different as regards the other question, that of the cost of "capital". At first sight, Keynes' position seems to coincide with that of the traditional theory, but, when we think about it, it seems clear that this is not so. The fact is that, in the traditional theory, the concept of marginal productivity of capital plays an essential role. This is not what we find in Keynes. The concept of "marginal efficiency of capital" which we find in the General Theory (Chapter XI) is radically different from the concept of marginal productivity of capital. This, as we have pointed out, must be regarded as the partial derivative of production in relation to a very slight change of one factor, assuming that the others are constant. The marginal efficiency of capital is formed of the series of future receipts that the entrepreneur expects to obtain from the sale of the product, after deducting the running expenses (a qualification similar to the one introduced by Marshall in defining the net marginal product in order to take account of the complementarity links between the factors which cannot be eliminated). To be more precise, the marginal efficiency of capital is the discount rate at which the present value of the series of receipts expected from the capital good during its life equals the supply price of that good. It follows that a fall in the market rate of interest leads to a greater purchase of capital goods, that is, causes the volume of investment to rise. It is here, in the consequences of a fall in interest, that there clearly emerges the radical difference between the two concepts. From the traditional point of view, such a diminution causes a change in the combination of techniques — capital intensity increases, labour intensity falls. From the Keynesian point of view, on the contrary, investment increases: the combination of capital and labour may change or remain constant — the diminution of interest does not permit us to arrive at any conclusion as to the combination of factors and on their substitution.

As is well known, Keynes, in the same rejoinder to Dunlop and Tarshis, expresses a highly positive judgement of the article which Kalecki had published in the April 1936 number of Econometrika to explain income distribution starting from the concept of mark-up, which he, following the traditional theory, conceives of as a "degree of
monopoly" (a view which was at the basis of quite a few sterile discussions). Kalecki, writes Keynes (p. 49), in that article "employs a highly original technique of analysis in the distributional problem between the factors of production in conditions of imperfect competition, which may prove to be an important piece of pioneer work".

It is not clear whether Keynes fully realized that Kalecki's explicative hypothesis, which he praises so highly, represented a radical departure from the marginalist model of distribution. But it does not appear that Keynes really accepted that model.

11. Concluding observations

The adjective "marginal" has therefore many meanings, some of them acceptable and others not. In general, the adjective is used correctly when it indicates a derivative which in its turn is the consequence of the application to economic problems of the maximum and minimum criteria derived from the infinitesimal calculus — an application which began to take a systematic turn slightly more than a century ago. Now it is true that, as a rule, the theoretical models to which I refer were elaborated in static or instantaneous terms — abstracting from time — and this strongly limits their interpretative value. But this approach is not at all inevitable in treating economic problems. In any case, if they are logically consistent and not based on elements obviously clashing with the reality which is being explained, these models are fully entitled to a place in economic theory. The criticisms as regards marginalism must be directed at the merit and not at the method.

This does not mean, however, that there are only formal differences between the marginalists and the critics of marginalism. In this article, I have dealt with some fundamental questions regarding the theory of the firm and the connected theory of prices and the theory of capital. In the conception of these questions, there are substantial differences. Recently, certain economists who defend the marginalist theory, including Frank Hahn, have asserted that, if polemical prejudices are set aside, the differences are more apparent than real. These economists, however, admit the validity of the criticism of the aggregate production function. Hence in fact, they abandon the marginalist theory of distribution, since they have not presented any alternative model. If this is so, then the marginalist construction is deprived of one of its major planks. The fact that certain particular models are still standing — to give a single example, the model for monopoly — has scant importance for the construction considered in its entirety. Indeed, for such a construction, it has no meaning.

To eliminate the ambiguity now ruling the roost on the theoretical scene, this is my thesis — we must have a critical reconsideration which makes no concession to the "let sleeping dogs lie" bent of intellectual life. In the case of diminishing returns and increasing marginal costs, many economists, in order to retain their attachment to traditional theory despite the impressive volume of theoretical and empirical analyses which show constant marginal costs and constant or increasing returns, have taken refuge in the idea of the "relevant" or visible section of the curves of marginal costs. This section, they say, is "almost constant", while the curve starts to climb, as the traditional theory demands that it should, as soon as it becomes invisible. In the case of the "cost of capital", the coexistence between the correct and the wrong view is justified in a somewhat different manner. Since it is impossible to deny that interest is the cost of loans, it is argued that this is obviously true in the first instance, but that the "cost of capital" must be related to capital taken as a whole, but physically determined — fixed and working capital, excluding labour. Many years ago, I noted this coexistence in no less an economist than Joseph Schumpeter who, in his Theory of Economic Development (1934), adopting the approach of the classics, with original and by no means superficial arguments, supports the validity of the point of view of businessmen for whom capital is a fund of purchasing power, while, in Business Cycles, he reasons as if the traditional theory were valid. In my book, The Forces of Economic Growth and Decline, I recalled the circumstances in which I contested this incredible coexistence of opposite ideas in Schumpeter's work (pp. 113 and 114 and note). If the two concepts could coexist in the mind of an economist of Schumpeter's standing — who was, intro
the bargain, an economist who, on the specific question of interest and the cost of capital, had put forward a relatively original and essentially correct view — we need hardly be surprised if this coexistence turns up again systematically among economists brought up on the traditional theory.

I feel it is time to leave such ambiguity behind us.

Roma

PAOLO SYLOS LABINI

REFERENCES


