Nicholas Kaldor

The Economic Journal, Vol. 82, No. 328, December 1972

THE IRRELEVANCE OF EQUILIBRIUM ECONOMICS[1]

THE purpose of my lecture today is to explain why, in my view, the prevailing theory of value—what I called, in a shorthand way, " equilibrium economics "—is barren and irrelevant as an apparatus of thought to deal with the manner of operation of economic forces, or as an instrument for non-trivial predictions concerning the effects of economic changes, whether induced by political action or by other causes. I should go further and say that the powerful attraction of the habits of thought engendered by " equilibrium economics " has become a major obstacle to the development of economics as a *science*—meaning by the term " science " a body of theorems based on assumptions that are *empirically* derived (from observations) and which embody hypotheses that are capable of verification both in regard to the assumptions and the predictions.

The word " equilibrium " in economics is used, of course, in all kinds of contexts—in Keynesian economics for example, or in theory of the balance of payments, and so on. I should therefore make clear that the notion of equilibrium to which I refer is that of the general economic equilibrium originally formulated by Walras, and developed, with ever-increasing elegance, exactness, and logical precision by the mathematical economists of our own generation, of whom perhaps the French economist, Gerard Debreu is now regarded as the most prominent exponent.[2]

Taken at its purest and most abstract level, the pretensions of this equilibrium theory are modest enough. Although Debreu describes the subject-matter of his book as " the *explanation* of the price of commodities resulting from the interaction of the agents of a private ownership economy,"[3] it is clear that the term " explanation " is not used in the ordinary everyday sense of the term. It is intended in a purely logical and not in a " scientific " sense; in the strict sense, as Debreu says, the theory is " logically entirely disconnected from its interpretation." It is not put forward as an explanation of how the actual prices of commodities are determined in particular economies or in the world economy as a whole. By the term " explanation " Debreu means a set of theorems that are *logically* deducible from precisely formulated assumptions; and the purpose of the exercise is to find the minimum " basic assumptions " necessary for establishing the existence of an " equilibrium " set of prices (and output/input matrixes) that is (*a*) unique, (*b*) stable, (*c*) satisfies the conditions of Pareto optimality. The whole progress of mathematical economics in the last thirty to fifty

[1] The Goodricke Lecture delivered in the University of York, May 10, 1972.

[2] *Theory of Value, An Axiomatic Analysis of Economic Equilibrium*, Cowles Foundation Monograph No. 17, New York, 1959.

[3] *Ibid.*, p. vii, italics mine.

years lay in clarifying the minimum requirements in terms of " basic assumptions " more precisely: without any attempt at verifying the realism of those assumptions, and without any investigation of whether the resulting theory of " equilibrium prices " has any explanatory power or relevance in relation to actual prices.

I. Axiomatic Theory and Scientific Hypothesis

It would take me too long to enumerate all these basic assumptions; it would also lead me away from my main argument. But unlike any scientific theory, where the basic assumptions are chosen on the basis of direct observation of the phenomena the behaviour of which forms the subject-matter of the theory, the basic assumptions of economic theory are either of a kind that are unverifiable—such as that producers " maximise " their profits or consumers " maximise " their utility—or of a kind which are directly contradicted by observation—for example, perfect competition, perfect divisibility, linear-homogenous and continuously differentiable production functions, wholly impersonal market relations, exclusive role of prices in information flows and perfect knowledge of all relevant prices by all agents and perfect foresight. There is also the requirement of a constant and unchanging set of *products* (goods) and of a constant and unchanging set of *processes of production* (or production functions) over time—though neither category, goods nor processes, is *operationally* defined: in other words, no attempt is made to show how these axiomatic concepts are to be defined or recognised in relation to empirical material.

While this pure theory is not *intended* to describe reality, it is put forward as the necessary conceptual framework—the necessary starting point—for any attempt at explaining how a " decentralised " system works; how individuals guided entirely by the market, or rather by price information, sort themselves out between different activities and thereby secure the maximum satisfaction both to themselves and, in the specific Pareto-sense, to society as a whole.

Indeed it is the deep underlying belief, common to all economists of the so-called " neo-classical " school, that general equilibrium theory is the one and only starting point for any logically consistent explanation of the behaviour of de-centralised economic systems. This belief sustained the theory despite the increasing (*not* diminishing) arbitrariness of its basic assumptions—which was forced upon its practitioners by the ever more precise cognition of the needs of logical consistency. In terms of gradually converting an " intellectual experiment " (to use Professor Kornai's phrase) [1] into a scientific theory—in other words, into a set of theorems directly

[1] J. Kornai, *Anti-Equilibrium. On economic systems theory and the tasks of research.* North Holland Publishing Co., Amsterdam, 1971, p. 11.

related to observable phenomena [1]—the development of theoretical economics was one of continual *de*gress, not *pro*gress: the ship appears to be much further away from the shore now than it appeared to its originators in the nineteenth century. The latest theoretical models, which attempt to construct an equilibrium path through time with all prices for all periods fully determined at the start under the assumption that everyone foresees future prices correctly to eternity, require far more fundamental " relaxations " for their applicability than was thought to be involved in the original Walrasian scheme. The process of removing the " scaffolding," as the saying goes,—in other words of *relaxing* the unreal basic assumptions—has not yet started. Indeed, the scaffolding gets thicker and more impenetrable with every successive reformulation of the theory, with growing uncertainty as to whether there is a solid building underneath.

Yet the main lessons of these increasingly abstract and unreal theoretical constructions are also increasingly taken on trust—as if in the social sciences, unlike the natural sciences, the problem of verification could be passed over or simply ignored. It is generally taken for granted by the great majority of academic economists that the economy always approaches, or is near to, a state of " equilibrium "; that equilibrium, and hence the near-actual state of the world, provides goods and services to the maximum degree consistent with available resources; that there is full and efficient utilisation of every kind of " resource "; that the wage of every kind and quality of labour is a measure of the net contribution (per unit) of these varying kinds and qualities of labour to the total product; that the rate of profits reflects

[1] The difference between a scientific theory and an " axiomatic " theorem has been well put by Einstein:

> " Physics constitute a logical system of thought which is in a state of evolution, whose basis cannot be distilled, as it were, from experience by an inductive method, but can only be arrived at by free invention. The justification (truth content) of the system rests in the verification of the derived propositions by sense experiences."
>
> " The skeptic will say: ' it may well be true that this system of equations is reasonable from a logical standpoint. But it does not prove that it corresponds to nature '. You are right, dear skeptic. Experience alone can decide on truth."

A. Einstein, *Ideas and Opinions*, New York, 1960, pp. 322 and 355 (quoted by Kornai, *op. cit.*, pp. 9–10).

The difference mainly resides in this. In the case of physics, any fundamental re-consideration of the basic " axioms " of the system is the result of observations which could not be made consistent with existing hypotheses. Examples (chosen at random) are the observation that the amount of radiation emitted by Pitchblende was greater than could be accounted for by the absorption of sunlight; that a stream of light which passed through a glass and was directed at a mirror at some particular angle is not reflected by the mirror; or that there is a " reddening " of the spectrum observed in distant stars. In economics, observations which contradict the basic hypotheses of prevailing theory are generally ignored: the " theorist " and the " empiricist " operate in two isolated compartments and the challenge of anomalous observations is ignored by the theorist—as something that could be taken into account at the stage of " second approximation " without affecting the basic hypotheses. And where empirical material is brought into conjunction with a theoretical model, as in econometrics, the role of empirical estimation is to " illustrate " or to " decorate " the theory, not to provide support to the basic hypothesis (as for example, in the case of numerous studies purporting to estimate the coefficients of production functions).

the net advantage of substituting capital for labour in production, etc., etc.—all propositions which the *pure* mathematical economist has shown to be valid only on assumptions that are manifestly unreal—that is to say, directly contrary to experience and not just " abstract." In fact, equilibrium theory has reached the stage where the pure theorist has successfully (though perhaps inadvertently) demonstrated that the main implications of this theory cannot possibly hold in reality, but has not yet managed to pass his message down the line to the textbook writer and to the classroom.

Yet without a major act of demolition—without destroying the basic conceptual framework—it is impossible to make any real progress. There is, I am sure, a vague sense of dissatisfaction, open or suppressed, with the current state of economics among most members of the economics profession —as is evidenced, for example, by recent Presidential addresses to the Royal Economic Society and to section F of the British Association.[1] On the one hand it is increasingly recognised that abstract mathematical models lead nowhere. On the other hand it is also recognised that " econometrics " leads nowhere—the careful accumulation and sifting of statistics and the development of refined methods of statistical inference cannot make up for the lack of any basic understanding of how the actual economy works. Each year new fashions sweep the " politico-economic complex " only to disappear again with equal suddenness—who can now recollect the great revival of the quantity theory of money of three years ago, or the more recent belief that frequent fiscal adjustments, guided by the best forecasting techniques, can maintain the steady growth of the economy at its pre-determined growth potential, not to speak of the Phillips Curve? These sudden bursts of fashion are a sure sign of the " pre-scientific " stage, where any crazy idea can get a hearing simply because nothing is known with sufficient confidence to rule it out.

II. Where Economic Theory Went Wrong

The difficulty with a new start is to pinpoint the critical area where economic theory went astray. In my own view, it happened when the theory of value took over the centre of the stage—which meant focusing attention on the *allocative* functions of markets to the exclusion of their *creative* functions—as an instrument for transmitting impulses to economic change.

To locate the source of error with more precision, I would put it in the middle of the fourth chapter of Vol. I of the *Wealth of Nations*. The first three chapters are devoted to the principle of the Division of Labour. These explain that the larger the production, the lower real cost per unit tends to be, because the larger the production, the more efficient the modes of

[1] E. H. Phelps Brown, " The Underdevelopment of Economics;" G. D. N. Worswick, " Is Progress in Economic Science Possible?," Economic Journal, March 1972, pp. 9–20 and 73–86.

production that can be employed: the greater the specialisation and the sub-division into different processes. In the first chapter Smith gave numerous reasons for this basic law, beautifully illustrated by the example of pin-making. In the second chapter he explains the peculiarly human characteristic of the propensity to truck, barter and exchange one thing for another—" nobody ever saw a dog make a fair exchange of one bone for another with another dog "—which alone makes it possible to develop the division of labour through social co-operation. Indeed for Smith the existence of a " social economy " and the existence of increasing returns were closely related phenomena. And the third chapter, perhaps the most significant of them all, is devoted to the proposition " that the division of labour is limited by the extent of the market "—a theorem which Allyn Young, writing 150 years later, (in a paper to which I shall refer more extensively presently) regarded as " one of the most illuminating and fruitful generalisations which can be found anywhere in the whole literature of economics."

But in the following chapter, after discussing the need for money in a social economy, Smith suddenly gets fascinated by the distinction between money price, real price and exchange value, and from then on, hey presto, his interest gets bogged down in the question of how values and prices for products and factors are determined. One can trace a more or less continuous development of price theory from the subsequent chapters of Smith through Ricardo, Walras, Marshall, right up to Debreu and the most sophisticated of present-day Americans.

The basic assumption of this theory is constant costs, or constant returns to scale. With Smith and Ricardo, this was implicit in the very notion of the " natural price " determined solely by costs of production (irrespective of demand). With the neo-classical school—in any rigorous formulation of it—it was explicit in the assumption of homogenous and linear production functions which is one of the required " axioms " necessary to make the assumptions of perfect competition and profit-maximisation consistent with one another.[1] Though Marshall, through the notion of " external economies " and the use of the partial equilibrium technique thought he could accommodate both increasing and decreasing returns to scale within the same analytical framework—an attempt which was shown to be logically faulty in Piero Sraffa's famous 1926 article on the Laws of Returns [2]—the general equilibrium school (as distinct from Marshall) has always fully

[1] This of course, embraces the classical case of increasing costs of production (in terms of labour and capital) due to the fixity of supply of land, provided the fixed factor earns its due rent. It is *not* consistent however with diminishing returns to scale—when *all* factors are increased in the same proportion, and the product increases in less than the same proportion—due, *e.g.*, to " external diseconomies."

[2] " The Laws of Returns under Competitive Conditions," Economic Journal, December 1926, p. 535. To be fair, Sraffa's critique had more relevance to the " Marshallian school " at Cambridge (and particularly to Pigou) than to Marshall himself who always expressed considerable doubt about the applicability of the theory of " normal price " to the case of increasing returns. (See particularly Appendix H of the *Principles*.)

recognised the *absence* of increasing returns as one of the basic " axioms " of the system. As a result, the existence of increasing returns and its consequences for the whole framework of economic theory have been completely neglected.

III. The Dominating Role of Increasing Returns

Yet on an empirical level, nobody doubts that in any economic activity which involves the processing or transformation of basic materials—in other words, in industry—increasing returns dominate the picture for the very reasons given by Adam Smith in the first chapter of the Wealth of Nations: reasons that are fundamental to the nature of technological processes and not to any particular technology.[1] One aspect of this is that *plant costs* per unit of output necessarily decrease with size in any integrated process of operation—such as a steel plant, a chemical plant, an electricity generator or an oil tanker—simply on account of the three-dimensional nature of space.[2] Provided the technical problems of construction can be solved, an increase in size is bound to bring further cost reductions since capacity is bound to increase faster than construction cost.[3] In the last decade, for example, there have been very large increases in the size of generating stations, of oil tankers and of the " optimal " steel plants, and there appears to be no reason why this process should come to a halt.

Another aspect, to which Allyn Young attributed major importance, is the break-up of complex processes into a series of simple processes, " some of which at least lend themselves to the use of machinery." He argued that the extent to which capital is used in relation to labour is predominantly a matter of the scale of operations—the capital/labour ratio in production is a function of the extent of the market rather than of relative factor prices.[4]

[1] As Smith emphasised in the first chapter, the opportunities for enrichment through a greater division of labour are far more important in manufactures than in agriculture: " The most opulent nations, indeed, excel all their neighbours in agriculture as well as in manufactures; but they are commonly more distinguished by their superiority in the latter than the former."

[2] For a discussion of this cf. G. C. Hufbauer, *Synthetic Materials and the Theory of International Trade*, London, Duckworth 1966, pp. 46 ff. For a much earlier account of the same idea, cf. E. A. G. Robinson, *The Structure of Competitive Industry*, Cambridge, 1931, pp. 29–31.

[3] For example, the cost of construction of a cylinder (or a pipeline) may be assumed to vary with the size of the diameter, since $2r\pi$ will indicate the size of the surface to be covered per unit of length. The capacity of the cylinder will grow on the other hand as the square of the radius, $r^2\pi$. Since a larger cylinder will require a thicker steelplate, the material costs will increase more than in proportion, but the labour costs will increase less than in proportion. Assuming that labour and material costs *together* vary in linear proportion to r, and assuming that one wished to describe this relation in terms of a " production function " of the Cobb–Douglas type (*i.e.*, with a constant elasticity of substitution of unity) the sum of the coefficients of the function would add up exactly to 2. See also Appendix on " Indivisibilities and Increasing Returns," below.

[4] " It would be wasteful to make a hammer to drive a single nail; it would be better to use whatever awkward implement lies conveniently at hand. It would be wasteful to furnish a factory with an elaborate equipment of specially constructed jigs, lathes, drills, presses and conveyors to build a hundred automobiles; it would be better to rely mostly upon tools and machines of standard types, *so as to make a relatively larger use of directly-applied and a relatively smaller use of indirectly-applied*

Finally, there are the inventions and innovations induced by experience to which Adam Smith paid the main emphasis—what we now call " learning by doing, " or " dynamic economies of scale." The advance in scientific knowledge in physics or in the science of engineering in the laboratory cannot by itself secure the innumerable design improvements that result from the repeated application of particular engineering principles. The optimum design for the steam engine or for the diesel engine or the sewing machine has only been achieved after many years or decades of experience: that for the nuclear power plant is still far away. The gain in design through experience is even more important in the making of plant and equipment; hence the *annual* gain of productivity due to " embodied technical progress " will tend to be all the greater the larger the number of plants constructed per year.[1]

It was left to Allyn Young to explore the main implications of Adam Smith's theorem on the manner of operation of economic forces in his famous article on *Increasing Returns and Economic Progress*, originally given as a Presidential address to Section F of the British Association in 1928.[2] On re-reading this paper after a lapse of many years, I feel convinced that it was so many years ahead of its time that the progress of economic thought has passed it by despite the attention it received at the time of its original publication. Economists ceased to take any notice of it long before they were able to grasp its full revolutionary implications. This was partly because Young was a man of exceptional modesty who underplayed, rather than emphasised, the full implication of what he was saying; his manner of exposition is suggestive, rather than compelling, and at times (as for example in the Appendix attached to the paper) obscure. It was partly also because its importance as a basic criticism of general equilibrium theory could not be appreciated at a time when that theory itself was not properly understood.

The consequences of abandoning the axiom of " linearity " and assuming that, in general, the production of any one commodity, or any one group of

labour. Mr. Ford's methods would be absurdly uneconomical if his output were very small, and would be unprofitable even if his output were what many other manufacturers of automobiles would call large." *Ibid.*, p. 530, italics added.

[1] On all these aspects there is a rapidly growing volume of empirical evidence, which makes the neglect of increasing returns by the theoretical model-builders all the more surprising. Taking only more recent publications, there is, apart from the sources cited by Hufbauer, *op. cit.*, the Annex on " *Industrial Profiles* " to the *Manual of Industrial Project Analysis* issued by the O.E.C.D. Development Centre (the Manual, but *not* the Annex, was prepared by I. M. D. Little and J. A. Mirrlees) which shows very large scale-economies in every one of the 18 types of industrial activities, such as brick-making, sugar manufacture, meat packaging, ironfounding, etc. for which detailed estimates are given. C. F. Pratten found (*Economies of Scale in Manufacturing Industry* D.A.E. Occasional Paper No. 28, Cambridge University Press, 1971) that of 44 types of activities examined, the minimum efficient scale for a *single plant* is 100% or more of total U.K. output in 7 cases, and in the range of 25–80% in 10 other cases. (This does not take into account, of course, economies due to greater differentiation and subdivision of processes.)

[2] Economic Journal, December 1928, pp. 527–542.

commodities, is subject to increasing returns to scale, are very far-reaching. The first and most important casualty is the notion of " general equilibrium " as such. The very notion of " general equilibrium " carries the implication that it is legitimate to assume that the operation of economic forces is constrained by a set of exogenous variables which are " given " from the outside and stable over time. It assumes that economic forces operate in an environment that is " imposed " on the system in a sense other than being just a heritage of the past—one could almost say an environment which, in its most significant characteristics, is independent of history. These critical exogenous features of the " environment " include Pareto's " tastes and obstacles "—the preferences of individuals as consumers, the transformation functions of factors into the products and the supply of resources—at any rate of " ultimate resources "—which are thus transformed. The notion of general equilibrium also assumes that the nature of the functions and of the social institutions—in particular the markets—are such that any given constellation of such exogenous variables will inevitably lead the system, possibly through a succession of steps, to a state of rest characterised by unchanging prices and production patterns over time: in other words that whatever the initial situation, the system will converge on a *unique point* the exact nature of which, both as regards the price system and the output system, can be deduced from the " data." Continuous economic change on these assumptions can only be conceived as some kind of " moving equilibrium " through the postulate of an autonomous (and unexplained) time-rate of change in the exogenous variables of a kind that is consistent with " continuous equilibrium " through time—such as a given rate of shift per unit of time in the production function of the so-called " Harrod-neutral " type or in the supply of resources: an exogenous rate of growth in the labour force and/or in the rate of increase in " capital "—though the very meaning of the latter concept has given rise to insoluble problems.

IV. The Theorem of Endogenous and Cumulative Change

Once however we allow for increasing returns, the forces making for continuous changes are *endogenous*—" they are engendered from within the economic system " [1]—and the actual state of the economy during any one " period " cannot be predicted except as a result of the sequence of events in previous periods which led up to it. As Young put it, with increasing returns " change becomes progressive and propagates itself in a cumulative way." [2] Further, " no analysis of the forces making for economic equilibrium, forces which you might say are tangential at any moment of time, will serve to illumine this field, for movements away from equilibrium, departures from previous trends, are characteristic of it." [3]

The basic consideration underlying Young's analysis is surprisingly the

[1] Young, *op. cit.*, p. 530. [2] *Op. cit.*, p. 533. [3] *Op. cit.*, p. 528.

same as that underlying Say's Law. If one takes an all-inclusive view of the economic process, economic activity ultimately consists of the exchange of goods against goods; this means that every increase in the supply of commodities enlarges, *at least potentially*, the market for other commodities. (The qualification " potentially," as we shall see, is very important and distinguishes Young's views from that of Say or Mill.) Hence the " extent of the market " depends on the division of labour almost as much, according to Young, as the division of labour depends on the extent of the market; and [quoting Young again] " modified . . . in the light of this broader conception of the market, Adam Smith's dictum amounts to the theorem that the division of labour depends in large part upon the division of labour. *This is more than mere tautology*. It means that the counter forces which are continually defeating the forces which make for economic equilibrium are more pervasive and more deeply rooted than we commonly realise." [1]

Myrdal, writing twenty-five years later, called this the " principle of circular and cumulative causation." [2] But neither Young nor Myrdal expressed the consequences in the radical form stated by Hicks [3] who said that " unless we can suppose . . . that marginal costs generally increase with output at the point of equilibrium " . . . " the basis on which economic laws can be constructed is shorn away." The words " economic laws " and " at the point of equilibrium " are of course question-begging. The issue is whether such laws (and " economic equilibrium ") exist or not. In the scientific sense, the postulate of the existence of such " laws " is refuted if they can be logically shown to be valid only under assumptions that are contrary to observed phenomena.

The whole issue, as Young said, is whether an " equilibrium of costs and advantages " is a meaningful notion in the presence of increasing returns.[4] When every change in the use of resources—every reorganisation of productive activities—creates the opportunity for a further change *which would not have existed otherwise*, the notion of an " optimum " allocation of resources—when every particular resource makes as great or greater contribution to output in its actual use as in any alternative use—becomes a meaningless and contradictory notion: the pattern of the use of resources at any one time can be no more than a link in the chain of an unending sequence and the very distinction, vital to equilibrium economics, between resource-creation and resource-allocation loses its validity. The whole view of the economic process as a medium for the " allocation of scarce means between alternative uses " falls apart—except perhaps for the consideration of short-run problems, where the framework of social organisation and the distribution of the major part of available " resources," such as durable equipment and trained

[1] *Ibid.*, p. 533. My italics.
[2] *Economic Theory and Underdeveloped Regions*, London, Duckworth, 1957.
[3] *Value and Capital*, Oxford, 1939, pp. 88–9.
[4] *Op. cit.*, p. 535.

or educated labour, can be treated as given as a heritage of the past, and the effects of current decisions on future development are ignored.[1]

Young saw clearly that the combination of Say's Law with Adam Smith's theorem is not enough in itself to ensure that change is progressive and " propagates itself in a cumulative way." Something more is needed linking the effects of changes of production to demand: something that would ensure that an increase in supply emanating from any particular part of the economy has a stimulating effect, and not a depressing effect, on production in other parts. Given that factor, the process of economic development can be looked upon as the resultant of a continued process of interaction—one could almost say, of a chain-reaction—between demand increases which have been induced by increases in supply, and increases in supply which have been evoked by increases in demand. Lacking a theory of income generation such as was supplied by Keynes in the General Theory eight years later, he thought that the necessary additional condition to ensure a continued chain reaction is to be found in the nature of reciprocal demand and supply functions—in other words, in the elasticity of Marshallian " offer curves," when the " commodities exchanged are produced competitively, under conditions of increasing returns." According to Young, when the demand for each commodity is elastic, " in the special sense that a small increase in its supply will be attended by an increase in the amounts of other commodities which can be had in exchange for it " progress is bound to be cumulative for " under such conditions an increase in the supply of one commodity *is* an increase in the demand for other commodities, and it must be supposed that every increase in demand will evoke an increase in supply. The rate at which any one industry grows is conditioned by the rate at which other industries grow, but since the elasticity of demand and supply will differ for different products, some industries will grow faster than others. *Even with a stationary population and in the absence of new discoveries in pure and applied science* (as contrasted with such new ways of organising production and such new ' inventions ' as are merely adaptations of known ways of doing things, made practicable and economical by an enlarged scale of production) *there are no limits to the process of expansion except the limits beyond which demand is not elastic and returns do not increase.*" [2]

[1] The only respect in which market prices have an indispensable " allocative " function to fulfil is that involved in the distribution over time of the use of *exhaustible* natural resources (*i.e.*, in the decision how far the current use of such resources should be restricted for the sake of the future) and it is notorious that it is in this respect that the price mechanism fails completely in making any allowance for the probable higher scarcity of such resources in the future.

[2] Young, *op. cit.*, p. 534, italics mine. In a footnote attached to the beginning of the above passage, Young also says that " if the circumstance that commodity *a* is produced under conditions of increasing return is taken into account as a factor in the elasticity of demand for *b* in terms of *a*, elasticity of demand and elasticity of supply may be looked upon as different ways of expressing a single functional relation." This almost suggests the view that the elasticity of demand for some commodities is a reflection of the elasticity of supply of other commodities.

V. The Role of Demand and the Two Kinds of "Induced Investment"

If the above passage has not received the attention which it deserved, it was, I believe, mainly because of the obscurity surrounding the meaning of "elasticity of demand" in the particular context. Clearly what Young intuitively perceived was that the pre-condition of cumulative change is that the rise in production of any one commodity *a*, should be associated with an increase in demand for all other commodities. He thought that this condition will be satisfied when the elasticity of demand for commodity *a* is greater than unity, since in that case the sales-receipts (or income) of the producers of *a* will be the greater the larger the production.

A little reflection will show however that if by "elasticity of demand" we mean something which is a reflection of the elasticity of substitution of consumers—in other words, of the elasticity of "flow" demand, as defined below—the increase in purchasing power of the producers of commodity *a* following upon the rise in the production of *a* must have been the result of a *diversion* of expenditure in favour of *a* and against other commodities. The rise in incomes of the *a* producers must therefore be offset by reduced incomes of the producers of some other commodities. It is possible that if the elasticities of substitution are high, and income elasticities are all positive, the elasticities of demand for all commodities, *taken individually*, should be greater than unity. But this is not enough to produce a chain reaction of rising demand followed by rising production, followed by rising demand, and so on, unless *total income* measured in terms of money is rising as well, which in turn presupposes that *total* expenditure, and not just the expenditure on a particular commodity, rises in response to a rise in production.

In order to show how an increase in the production of a commodity may involve the generation of additional incomes which in turn generates additional demand for other commodities and thereby becomes a "chain" in a continuous sequence, we must first of all take into account the fact that there are two kinds of demand (and supply) in a market: a "flow" demand and a "stock" demand: the former is the demand and supply of "outsiders," (*i.e.*, producers and consumers), whereas the latter represents the demand (or supply) originating from *inside* the market.

In pure theory the existence of this "stock" demand or "inside demand" is ignored. In a state of equilibrium, production and consumption, or "flow"-demand and "flow"-supply, are necessarily equal in each market, and in the rarefied world of Walrasian perfection where markets are *continually* in equilibrium, the question of how the market responds to "disequilibria" does not arise because all such "disequilibria" are ruled out—all equilibrating adjustments are assumed to be instantaneous, either because changes are timeless or because all changes have been perfectly foreseen.

However, the markets of the real world are not in continuous equilibrium in this sense; there are, or can be, persistent differences between production

and consumption which are reflected in increments or decrements in stocks. The impact effect of any undesigned or unexpected rise in production (due to a bumper harvest, for example) must be a rise in stocks; any subsequent adjustment in flow-demand or supply due to consequential price-changes requires time to materialise. For that reason, competitive markets are inconceivable without intermediaries—merchants or " dealers "—who are both buyers and sellers at the same time (at different prices) and who carry stocks so as to make " a market " that enables producers to sell and consumers to buy.

The size of the difference between their buying and selling prices (normally called the " dealer's margin ") depends both on the degree of perfection the market in which they operate and on the amount of " processing " or " transformation " performed by them. This may consist of pure merchanting activities—such as transportation, breaking bulk, packaging, etc.—and could also include varying degrees of physical transformation through manufacture. But what differentiates a merchant from other economic agents (such as a " producer ") is that his natural response to " outside " influences is to vary the size of his stock—to absorb stocks in the face of excess supplies and to release stocks in the face of excess demand. The merchants' function in other words is to create and preserve an " orderly " market which they can only do through their willingness to act as a shock-absorber: through their readiness to enlarge their commitments when prices are sagging and to curtail commitments when they are rising. The very notion of " merchanting " or " commercial " activities involves therefore the assumption that there is a certain elasticity of demand for *holding* stocks by the traders: an elasticity which is ultimately governed by the traders' expectations concerning prices and selling opportunities in the future. In a paper published many years ago I called this factor the " elasticity of speculative stocks " in a market [1] though the term " speculative " was perhaps a misnomer. It is true of course that traders only carry stocks in the expectation of making a profit, and therefore any inter-temporal transfer of goods could be called a form of " speculation," though it is fundamentally no different from any geographical transfer; and since the transportation of goods takes time, merchanting activities normally involve transfers of both kinds.[2]

[1] " Speculation and Economic Stability," *Review of Economic Studies*, October, 1939, p. 7 (reprinted in *Essays on Economic Stability and Growth*, London, 1960, p. 30).

[2] Any kind of merchanting activity—buying things with a view to their subsequent re-sale—is " speculative " in the sense that it involves the assumption of risks: by carrying stocks, traders *deliberately* take an " open position." Hence an increase in investment in stocks which occurs in response to an increase in supplies, though " induced," is a form of " voluntary " investment and not " involuntary." On the other hand, an increase in stocks which occurs as a result of a disappointment in sales-expectations—failure to *close* a position at the time and to the extent expected—may be regarded as " involuntary investment," in the sense that the addition to stocks *ex-post* must have been greater than that planned *ex ante*. However, any step which implies an increase in commitments—the " opening " of a position—may be assumed to be deliberate, even when in response to events which may have been unforeseen.

It is a hen-and-egg question whether historically it was the growth of commerce which continually enlarged " the size of the market " and thereby enabled increasing returns to be realised, or whether it was the improvement of techniques of production and the improvement in communication which led to the growth of commerce. In the process of the development of capitalism the two operated side by side. And it involved a tendency for a continual rise in the *value* (and not just the volume) of stock carried by traders in the markets, which meant in turn that the growth of production resulting from any favourable change on the supply side led to a growth in incomes which in turn generated an increase in effective demand for commodities.

The essential element missing from Young's presentation, and which can only be supplied on the basis of Keynesian economics, is the addition to incomes resulting from the accumulation of capital (in other words, from investment expenditure) combined with the induced character of such investment which arises more or less as a by-product of changes in the organisation of production.[1] It operates moreover in two different ways. In the really " competitive " markets, such as those for most primary products, which approximate the economist's notion of perfect competition (where individual buyers and sellers are faced with infinitely elastic demand and supply curves, and where increasing returns *cannot* be operative, at any rate at the level of the individual producer) the stocks which are essential for the functioning of the market are carried by merchants who are independent both from the producers and the consumers; it is their ability to act as a buffer—to absorb stocks in the face of a short-term excess of supply and *vice-versa*—which will lead to induced investment in the face of a rise in production: provided that the merchants' expectation of future prices make it appear profitable for them to increase the *value* of their stocks (and not only their volume) when prices sag in the face of excess supply.[2] In the markets for commodities in which increasing returns are important, and which, for that very reason, are only " imperfectly " competitive—as is the

[1] On re-reading Young in the light of Keynes, one is tempted to quote Keynes' account of Marshall's view that " . . . those individuals who are endowed with a special genius for the subject and have a powerful economic intuition will often be more right in their conclusions and implicit presumptions than in their explanation and explicit statements. That is to say, their intuitions will be in advance of their analysis and their terminology. Great respect, therefore, is due to their general scheme of thought, and it is a poor thing to pester their memories with criticism which is really verbal." (ECONOMIC JOURNAL, September 1924, p. 235, note, reprinted in *Essays in Biography*, p. 232.)

[2] Strictly speaking, there should be a net demand effect in real terms whenever there is an increase in the volume of stocks carried in relation to turnover (and not only when there is a rise in the total value of stocks) since any such increment implies a rise in investment (in real terms) in relation to output. However, when the merchants' elasticity of demand for increasing the stock-turnover ratio is less than unity—so that a 1% increase in the volume of stocks carried requires more than a 1% reduction in price—the purchasing power of producers will diminish in consequence of a rise in production, while the (theoretically) more-than-offsetting rise in the real purchasing power of consumers will be slow to percolate through the system.

case with manufactures—the producers carry their *own* stocks and adjust the rate of their production in response to changes in their sales (or in the state of their " order book ") and there will be " induced investment " in response to an *increase* in demand and the associated depletion of stocks. Such induced investment will partly take the form of circulating capital—that is to say, of an increase in the value of goods in process that is inevitably associated with the rise in production—and partly of fixed capital, in so far as the rise in current sales causes a revision of expectation of future sales.

It may seem paradoxical that " induced investment " should result from both increases in supply and increases in demand, but there is nothing necessarily inconsistent in this, provided there is asymmetry in market organisation between the two kinds of commodities, primary products and manufactures, an asymmetry which is imposed on the system by the differing incidence of the theorem of the " division of labour " between industry and agriculture—a feature of life which was already noted by Adam Smith. If, in the first approximation, one regards the essential division in economic activities as that between manufacturing activities and land-based activities (agriculture and mining) which provide the inputs (the food and raw materials) for manufacturing activities, and if we suppose that the quasi-automatic process of growing diversification and technological improvement resulting from the growth of activities—in other words increasing returns in the broad sense—is mainly a feature of the latter rather than the former, then the process of endogenous self-sustained growth requires both a certain *inelasticity* of expectations concerning *prices* (in regard to primary products) and also a certain *elasticity* of expectations concerning the *volume* of sales (in regard to manufactures). Induced investment reflecting the " acceleration principle " is a property of the latter; induced investment reflecting the price-stabilising effect of the operation of traders is a property of the former.[1]

And it requires, above all, a monetary and banking system that enables capital investment to increase in response to inducements, so as to generate the savings required to finance additional investment out of the *addition* to production and incomes. This is the real significance of the invention of paper money and of credit creation through the banking system. It provided the pre-condition of self-sustained growth. With a purely metallic currency, where the supply of money is given irrespective of the demand for credit, the ability of the system to expand in response to profit opportunities is far more narrowly confined.

[1] In post-Keynesian models of cycles and growth (such as, for example, Hicks, *A Theory of the Trade Cycle*, Oxford, 1950) the only kind of " induced investment " considered was the demand-induced kind—the kind relevant to the manufacturing sector. The other, induced by excess supply, was completely neglected.

VI. Some Conclusions

To end, we can do little more than to sketch some of the main consequences of this marriage of the Smith–Young doctrine on increasing returns with the Keynesian doctrine of effective demand. I should like to make three observations.

First, the sharp distinction made by Keynes between a " full employment " situation where real income is confined by resource-endowment, and an unemployment situation where it is limited by effective demand, disappears in the presence of increasing returns. Except in a purely short-term sense, total output can never be *confined* by resources. At any one time, there is, or there may be, a maximum potential output for the world as a whole resulting from past history which has determined the existing network of institutions and organisations, the different kinds of plant and equipment available and their geographical distribution, as well as the distribution of the available labour in all the different areas and their educational endowments and skills. Over a period, there may be a maximum *rate of growth* of output determined by the maximum rate of growth of production in some key sectors of the economy (such as the food-producing sectors) which limits the sustainable rates of growth of the other sectors. If that happens, it must be on account of the scarcity of natural resources, and the impossibility of substituting capital goods for natural resources at more than a certain speed, on account of an insufficiency of land-saving innovations. But if we take an inclusive view, neither labour nor capital can limit either the level, or the rate of growth, of production over a longer period. Capital accumulation can always be speeded up—or rather it automatically *gets* speeded up, with a faster growth of production. In the case of labour, there is no such thing as an " optimal " distribution of the labour force—with each man making a greater contribution to output in his existing employment than in any alternative employment—since every re-organisation of production resulting from overall expansion or new investment will mean the transfer of some of the labour force to new employments where its contribution to production will be greater than before. Just as Young emphasised that the adoption of more roundabout methods of production, due to an increase in the size of the market, and the adoption of more capital-intensive processes, are different facets of the same thing, so in the case of labour, no valid distinction can be made between an increase in the effective labour supply due to a rise in numbers employed and that due to a rise in productivity secured by a re-deployment of labour.

Second, it is evident that the co-existence of increasing returns and competition—emphasised by Young and also by Marx, but wholly excluded by the axiomatic framework of Walrasian economics—is a very prominent feature of de-centralised economic systems but the manner of functioning of which is still a largely uncharted territory for the economist. We have

no clear idea of *how* competition works in circumstances where each producer faces a limited market as regards *sales* and yet a highly competitive market as regards *price*.

Third, it is evident from our analysis that the " self-sustained growth " of decentralised economic systems, largely directed, not by exogeneous factors, but by the growth and the constellation of demand, is a fragile thing which will only proceed in a satisfactory manner if a number of favourable factors are present simultaneously: such as merchants who are ready to absorb stocks in the short run rather than allow prices to fall too far—because experience has taught them that market prices have some long-run stability —and manufacturers who respond to the stimulus of growing sales with an expansion of productive capacity, because experience has taught them that over a period markets are growing and not stable. It also requires a " passive " monetary and banking system which allows the money supply to grow in automatic response to an increased demand for credit.[1]

In the nineteenth century, with the background of rapid technological change, particularly in transport and communications, all these factors seem to have been present. In the present century, continued growth seems to have owed more to active government intervention—in the primary producing areas, through government-operated buffer stocks for commodities; in the industrialised countries, through " Keynesian " fiscal policies; both of which secured the continued growth of *real* purchasing power (*i.e.*, of effective demand in real terms and not just in money terms) without which economic growth would quickly grind to a halt.

NICHOLAS KALDOR

King's College,
Cambridge.

Date of receipt of final typescript: June 1972

APPENDIX

ON INDIVISIBILITIES AND INCREASING RETURNS

In an article published in this JOURNAL in 1934 I wrote that

> " . . . it appears methodologically convenient to treat all cases of large-scale economies under the heading " indivisibility." This introduces a certain unity into analysis and makes possible at the same time a clarification of the relationships between the different kinds of economies. Even the cases of increasing returns where a more-than-proportionate increase in output occurs merely on account of an increase in the amounts of the factors used, without any change in the proportions of the factors, are due to indivisibilities; only in this case it is not so much the " original factors," but the specialised functions of those factors, which are indivisible." [2]

[1] In a forthcoming paper on " Growth and Fluctuations in the World Economy," T. F. Cripps examines these conditions in a more systematic framework.

[2] " The Equilibrium of the Firm," ECONOMIC JOURNAL, March 1934, p. 65, reprinted in *Essays on Value and Distribution*, Duckworth, 1960, p. 39.

This proposition was later criticised in some detail by E. H. Chamberlin [1] while my own view was subsequently defended by Tjalling C. Koopmans.[2] I did not participate in this subsequent controversy since the question of whether increasing returns are " fundamentally " due to indivisibilities or not, did not then appear to me a matter of great moment. Recently, however, on reading Professor Koopmans' defence of my 1934 views, I have come to the conclusion that I ought to make a belated apology to the memory of the late Professor Chamberlin and acknowledge that he was basically right in his main contention—even though I was not persuaded by his arguments at the time.

The point is of more than semantic interest since if indivisibilities were the sole cause of increasing returns, there would always be some level of production at which such scale economies were exhausted and " optimum scale " production reached. Moreover, the prevalence of competition could itself be taken as an indication that the effects of " indivisibilities " are not such as to prevent optimum-scale production prevailing for a sufficiently small fraction of total output to be consistent with a reasonable approximation to perfect competition.

As was shown above, not all causes of increasing returns can be attributed to indivisibility of one kind or another and there is no reason to suppose that " economies of scale " become inoperative above certain levels of production. There is first of all the steady and step-wise improvement in knowledge gained from experience—the so-called " dynamic economies of scale " which have nothing to do with indivisibilities. But even in the field of " static " or " reversible " economies, there is the important group of cases which I described above as being due to the three-dimensional nature of space—*i.e.*, the fact that the capacity of, say, a pipeline can be quadrupled by doubling its diameter while the costs (in terms of labour and materials) are more nearly related to the diameter than to its capacity. There is nothing " indivisible " about tubes or pipelines as such: technically, it may be just as easy to make tubes of a relatively small or a relatively large dimension and there can be a continuous range of sizes in-between; the existence of a non-linear relationship between costs and capacity is inherent in the nature of *space*, and there is nothing " indivisible " about space as such. Moreover, this " space principle " applies equally to non-durable items (like plastic containers or paper bags) no less than to durable equipment (like steel pipes).

Professor Koopmans mentions the case of the pipeline explicitly but misses the point of the example:

> " I have not found one example of increasing returns to scale in which there is not some indivisible commodity in the surrounding circumstances. The oft-quoted case of a pipeline *whose diameter is continuously variable* can be seen as a case of choice between alternative pieces of capital equipment, differing in diameter, used to carry oil from Tulsa to Chicago, say. *No matter what diameter is selected, one entire pipeline of the requisite length* is needed to render this service. Half the length of the line does not carry half the flow of oil from Tulsa to Chicago." [3]

[1] " Proportionality, Divisibility and Economies of Scale," *Quarterly Journal of Economics*, February 1948, reprinted as Appendix B to the sixth edition of *The Theory of Monopolistic Competition*, Cambridge, Mass., 1948.

[2] *Three Essays on the State of Economic Science*, New York, McGraw Hill, 1957, pp. 150–2.

[3] *Op. cit.*, p. 152, n. 3. Italics mine.

There is a clear misunderstanding here as to the relevance of the indivisibilities involved to the existence of increasing returns. This has nothing to do with the *length* of the pipeline but only with the *width* of the pipeline: the " indivisibility " on the other hand, (as Koopmans says) relates to the length and not the width. Increasing returns arise because the capacity of a pipeline of *unit length* to carry oil—*i.e.*, the maximum volume of throughput per unit of time—increases with the square of diameter, whereas the cost of production is a linear function of it. If a pipe of 5 feet diameter can transmit 5,000 tons per hour, a pipe with 10 feet diameter will transmit 20,000 tons per hour, and so on.

Professor Koopmans' method, if I understand him correctly, is to treat pipelines of differing diameters as different " commodities," so that the choice of a pipeline with a particular diameter comes to the same as the choice of a particular " linear activity," or process of production. He regards every produced commodity which has the characteristic that " the ratios of inputs into their manufacture to outputs from their use cannot be reproduced at a smaller scale " as undergoing a " qualitative change " with every change in the ratio of inputs to outputs.[1]

However, each of these " linear processes " would only be relevant for a *particular* output,[2] and there is also an underlying functional relationship between outputs and inputs which may show perfect continuity but which is basically non-linear. This underlying relationship links the quantity of oil transmitted per hour as the " output " and the labour, materials, etc., involved in constructing the pipeline and all other associated outlays as the " inputs."

Professor Koopmans agrees with Chamberlin that his definition of " commodities " makes the whole issue a tautological one [3] but he believes that nevertheless the indivisibility has the right " intuitive connotations ":

> " . . . the reproach of tautology has been levelled against many propositions of economic theory. What matters is that a model which differs from the linear activity analysis model in that it omits the proportionality postulate or at least excepts from it all activities involving certain commodities seems to express those aspects of reality that have been recognised as responsible for increasing returns to scale. Such a model may therefore be a suitable vehicle for a first exploration of this phenomenon, and on the suitability of prices as guides to allocation. So far, mathematical difficulties have been the main obstacle to such an exploration." [4]

The significance of all this depends on what is meant by the " suppression of the proportionality postulate." At one end, it may mean nothing more than the introduction of discontinuities which may rob the analysis of some of its elegance and simplicity, but without destroying the existence of a convex " Pareto-frontier " of some kind. At the other end, it may mean that the whole notion of a Pareto-optimal equilibrium and of the price mechanism as a means of bringing about an " optimal " resource-allocation becomes illegitimate.

[1] *Op. cit.*, pp. 151–2.

[2] Defined in this case as the *throughput* of oil (at any particular point) per unit of time.

[3] Although the statement, quoted earlier, that he has not "found a single case in which there is not some indivisible commodity in the surrounding circumstances " would suggest that he regarded the proposition as a factual one, and not as a logical (tautological) one.

[4] *Ibid.*, p. 152.

Allowance for indivisibilities means that for activities involving certain commodities there is a minimum scale of output, and the activity can only be " attained " at integral multiples of that minimum scale. If in an actual economy the level of output of any one final commodity is some multiple of the minimum output of the " best " available technique for producing it, the existence of indivisibilities will simply mean that the " efficiency frontier " becomes a " jagged surface " instead of a smooth one, but yet remains convex in the large.

However if at any actual level of output the " best " available technique for that output is less efficient than that available for a somewhat larger output—if, in other words, there is a whole hierarchy of activities not all of which are feasible or attainable at any point of time—the choice among " activities " becomes primarily a matter not of prices but of the scale of production. With every enlargement of production new " activities " become profitable which could not have been employed earlier, whilst the introduction of such new " activities " leads to the invention of further " activities " which have not been " known " earlier.

Since (as was argued above) the demand for any particular product or group of products is a reflection of the level of production of other products, this means that any re-allocation of resources which enlarges the range of feasible activities comes to the same as an " outward shift " in the production frontier. The problem then becomes not just one of " solving the mathematical difficulties " resulting from discontinuities but the much broader one of replacing the " equilibrium approach " with some, as yet unexplored, alternative that makes use of a different conceptual framework.