Introduction. Cracks in the neoclassical mirror: on the break-up of a vision

Edward J. Nell

By now, most economists are aware of the flaws in the aggregative neoclassical models of growth, productivity, and income distribution. Whether neoclassicism in its full generality is open to the same sorts of objections is still a matter of dispute. It is not the object of this book to enter that dispute. Instead, it will first chart some of the territory that neoclassicism has been forced to yield, and then stake out a claim and begin building new edifices there. This introduction will try to relate one of the central purposes of these constructions – understanding what capital is – to the critique of neoclassical theory.

For many economists neoclassicism is economic science, and attacks upon it simply create a sense of unease. Without theory, there is no science; economists will be left with nothing to say that could not be said as well by any intelligent observer who takes the trouble to study the facts. But the attacks on neoclassicism are not attacks on theory as such. They are attacks on a theory regarded as wrong and have been designed as a prelude to replacing that theory with a better one. As will become evident, there is more agreement on the defects of orthodox theory than there is on what theory is to replace it; but all are agreed that the point of the criticism is to clear the ground for construction.

The standard version of orthodox theory

The critique of neoclassicism grew out of a concern with the way orthodox economics treated capital and wage–labor, and has been principally directed against what may be called the standard version of orthodox theory. This latter is not necessarily a “one-sector” or an aggregate model, although the most favored presentation of the theory normally takes that form. What distinguishes it, rather, is the way the supply side is treated. All forms of neoclassicism postulate households where preferences are described by utility functions, as the basis for demand. The standard version treats supply symmetrically – postulating the existence of a definite number of firms each of whose technical production possibilities are described by a production function relating factor inputs and outputs. However, neoclassicism can also be developed using the methods of activity analysis, in which case no firms are represented at all. Production functions in activity analysis show the inputs required for a product, rather than the inputs required by a firm. Only firms can make market decisions; the cost of a product, in the abstract, may be interesting, but it is the cost to the firm, not in the abstract, that will be relevant to the firm’s strategy in the marketplace. The standard version is a theory of the marketplace, and tries to deal with and classify different market forms, according to the competitive environments firms face and create for one another.

It is because the standard version deals with the concrete activities of firms, rather than with the abstract technology of products, that it must employ what is misleadingly referred to as an “aggregative” concept of capital. The point is not so much that the various inputs under the firm’s control are valued and aggregated. As participants on all sides of the capital controversy have observed, this in itself is of little interest. Everyone agrees that given the prices, it is easy to aggregate capital goods; without prices no proxy measure will do. The question is, why bother aggregating capital?

The answer is as basic as it is obvious. A firm’s capital is what makes it what it is. The firm is the institutional form which a particular capital takes. Its permanent existence is not as a set of capital goods, but as a fund of capital. The fund will be embodied from time to time in capital goods, such as plant and equipment, and
Edward J. Nell

inventories. But capital remains while capital goods are used up, inventories are converted to output and sold off, all of which is another way of saying that the activity of a firm is to turn over its capital, making a profit in doing so. Marx captured the essence of this process in a simple formulation, encompassing what he called the three circuits of capital – the circuits of production, commodity, and money capital.

\[ M \xrightarrow{L} C \xrightarrow{P} M \]

\[ \ldots \]

\[ M \xrightarrow{L} C \xrightarrow{M'} \xrightarrow{c} m \]

Money capital is used to buy means of production and labor which then produce an expanded set of commodities, \( C' \), sold again for money, \( M' \), which then exchanges for productive commodities, \( C \), and for luxury consumption, \( c \) (Marx, 1967, Vol. II, Ch. 4).

Marx’s way of representing the circuits of capital is valuable because it highlights the fact that in turning over, capital is regularly and repeatedly transformed from capital goods into inventories, then sales revenue, and finally, into capital goods again. At the point when it is in money form, the managers of capital must decide on the most advantageous selection of capital goods. The inherent – and sequential – connection between capital as a fund and capital as plant and equipment is made plain. Of course on any realistic account there are restrictions on the extent to which changes in the form of invested capital can be made at any given time. But this is no comfort to those who assume “malleable capital.” To the extent that capital is malleable, and conditions competitive, there will be movements of capital towards the highest rate of return. J. B. Clark, who understood very well the dual modes of existence of capital as a fund and as instruments of production, gives the example of New England capital leaving whaling for textiles. No ships were converted to mills. “As the vessels were worn out, the part of their earnings that might have been used to build more vessels was actually used to build mills. The nautical form of the capital perished; but the capital survived and, as it were, migrated from the one set of material bodies to the other.” (Clark, 1893, p. 118). Such mobility has traditionally been taken to establish a tendency to form a uniform rate of profit. At the very least nonuniformity of the rate of profit on capital funds is inconsistent with equilibrium in competitive conditions.

An analogous point holds true for labor. Competitive workers are on short-term contracts – daily wages – that either side can abrogate costlessly. Hence workers will always seek employment at the highest wages, while employers will try to hire only those willing to work for the lowest wages. The only position consistent with equilibrium is one in which the wage is uniform, at a level which clears the labor market. (Of course wages will not be uniform as between noncompeting labor groups but within such groups they must be uniform, and must clear the market for each group.)

Such points are elementary, and are of course widely recognized in the standard version of neoclassical theory, which seeks to determine the rate of interest on capital, the wage rate of labor and the rent of land in the factor markets, “coordinating the laws of distribution” through the theory of marginal productivity. But, surprisingly, these elementary features of competition are not consistent with the activity analysis version of neoclassicism, at least as ordinarily presented.

Main features of neoclassicism

Activity analysis emphasizes different aspects of the market system, and in many ways marks a new departure in neoclassical thinking. It is analytically more powerful, and it permits a far more detailed representation of the economy. To understand the relation between activity analysis and the thinking that preceded it let us try to sum up the main features of neoclassicism in a few basic propositions. This will not be easy. Neoclassical writers had different interests, and frequently advanced different and competing theories. But there are certain common threads, and these can be woven into a fabric of ideas. Not all neoclassical writers would be content to wear the resulting garment, at least without alterations, and some might claim it was made up of whole cloth. Nevertheless the following ideas do constitute a complete, coherent and thoroughly familiar picture of a market economy, obviously recognizable as the common ground of most of the profession.

The point. The purpose of neoclassical economic theory is (a) to exhibit and explain the working of the market system, or the price system; thus to establish the forces determining prices and quantities exchanged, (b) to establish the circumstances under which and the extent to which the market system encourages efficiency in allo-
cation of resources, (c) to determine the effects of market conditions on the economic welfare of the agents, as measured by their preferences.

The analysis may be carried out for an economy with given resources and endowments, or for an economy with expanding resources.

The method. Neoclassical thinking proceeds by means of the behavioral equilibrium method. Prices, rates of return (usually treated as prices, e.g., interest is the price of capital) and sometimes quantities are signals, to which behavior and decision making respond. Supply and demand functions thus have a stimulus–response form. Equilibrium is defined in two ways: (a) Market clearing, the model is solved for that set of signals and responses consistent with clearing, (b) Pareto-optimal, the model is solved for that set of signals and responses that leaves everyone in the most preferred position attainable without pushing others to a less preferred position.

The two notions of equilibrium are closely linked but not identical. The method is to solve the model for equilibrium, and then predict that actual behavior will tend to approximate equilibrium behavior, or that it will be understandable as a deviation from equilibrium caused by an identifiable special circumstance.

An assumption. The method of behavioral equilibrium implies that there must be some tendency to equilibrium in some sense. Otherwise there would be no reason to solve the model for equilibrium – if it is no more likely in reality than any other position, why bother with it? (Unfortunately for the method, in some very important models it can be shown that no such tendency exists in general. Worse, unstable equilibria are the least likely points to be found in reality. This either requires modification of the model, or it means that the model can only illuminate reality by contrast.)

The agents. The agents whose behavior is to be studied can be separated into two broad groups. Agents who demand final products and supply factor services are households; agents who supply final products and demand factor services are firms. The same agent, playing different roles can appear in both groups. Households, behaving as demanders, exhibit their preferences in the product markets; firms, behaving as suppliers, exhibit a desire for earnings constrained by technological possibilities. Households and firms reappear in the factor markets, as suppliers exhibiting preferences such as work/leisure time, and demanders, reflecting the technological conditions which determine costs. The behavior of the agents is described by carefully specified functions, showing how far the agent should carry a desired course of actions, and stating exactly the constraints binding him. Objective factors, like initial endowments, are clearly separated from subjective ones, like expectations and uncertainty.

The markets. Markets can be divided mutually exclusively, though not exhaustively into markets for final products and markets for factors of production. Intermediate goods can be neglected – they swim around in Pigou’s lake (Hicks, 1946, p. 118; Pigou, 1932; Clark, 1893). Market analysis proceeds by building on the analysis of the behavior of the individual units, the firm, and households. The object is to determine the equilibria of these units, and then by aggregation, of a market as a whole. From this one proceeds to groups of related markets and finally to the general equilibrium of the whole system, product and factor markets both. Sometimes, as when pure exchange is considered, the analysis is confined to product markets. Clearly this is only an approximation. If a factor market is not in equilibrium, there will be shifts in the supply functions of any products employing that factor, as its market adjusts. For neoclassical theory production is a one way street running from factors, of which there are initial endowments, to final products. The markets for factors and for final products are markets in the same sense, and are to be analyzed in similar ways. However, the two groups of markets are interconnected. The demand for factors is derived from the demand for final products; the supply of factor services is the source of the income which makes product preferences effective.

Supply and demand. Equilibrium is arrived at through a balance of the forces of supply and demand, whether in one market or in many interrelated ones. Supply functions are based on choices from among given methods of production; demand functions are based on preferences for goods, and the whole is constrained by the initial endowments. The range of choice is assumed to be very wide so that systems with narrowly constrained choices are treated as special cases, which implies that nothing essential to the working of markets depends on the constraints on choice. Supply and demand are both necessary to the determination of both prices and quantities. (Thus in traditional welfare economics giving rise to both consumer and producer surpluses.)

Roles and class. Nothing essential depends on who or how many are capitalists and workers. Social classes can be introduced, but do not ap-
Edward J. Nell

pear in the most general models. Whether one is a capitalist or a worker is a matter of preference – labor can hire capital just as readily as capital can hire labor, in the absence of inflexibilities and imperfections in markets.

The preceding delineation of general neoclassical features is a broad enough approach to encompass both the standard version and the general equilibrium model of Walras and Cassel. The standard version incorporates the effect of competition in equalizing the wage on all substitutable grades and skills of labor, and in equalizing the rate of return across the entire field of capital goods within which capital funds can be invested. The Walras-Cassel activity analysis model treats each capital good and each type of labor separately, determining their efficient use and particular rental values.

Specifying the initial endowments

The distinction between the two senses of capital – funds versus factories – leads to a difficulty when we come to consider how neoclassical theory presents the social context within which economic activity is to take place. A universal property system is assumed, sometimes explicitly, more often implicitly; all means of production are owned, generally by private institutions, normally households or firms, where ownership implies alienability. Hence all products, including liabilities incurred in production are the property of the owner of the means of production, or of the contractual operator of those means, according to the terms of the contract. So far, so good; correct enough, though sparse in detail. The trouble comes in specifying the initial endowments, for neoclassical theory is not about to try to explain how property of various kinds came to be concentrated in certain hands. Given the initial distribution of ownership of factors, the theory determines equilibrium in produce and factor markets, and examines the conditions under which these equilibria define efficient allocations of resources and welfare maximizing patterns of consumption. But how the initial distribution of ownership came about is none of its concern.

However, it still has to assume that such an initial distribution exists, and here is where a problem arises for the standard version. Households’ initial endowments of land and labor are easily specified; the appropriate move would be to endow them with capital, also. But capital funds, or capital goods? If households are given specific machines, we are no longer operating with the broad concept of the factor. Households will trade their endowments of machines each trying to obtain the most profitable selection. More of this later; the point is that initial endowments of specific goods will be traded. So we have left the framework of the standard version. But if households are endowed with funds, we have to ask how such a value concept can be given meaning in advance of the determination of prices.

Even endowing households with funds, however, does not free the standard version from problems in its account of the factor market. There is a general difficulty concerned with the relationship between the concept of capital, as a factor supplied by households and demanded by firms, and the concept of the firm itself, the owner of the means of production and therefore of the product and the liabilities incurred in production. This difficulty suggests that capital may be more complex than capital theorists have realized. It is neither funds, as in the standard version, nor goods, as in activity analysis. Neither will it do to treat it as both together, goods and funds, at successive stages in circulation. There are other aspects to capital, not expressed by either goods or funds. Let us explore this further. To do this we must examine a problem that arises even in these very general accounts of neoclassicism.

A neoclassical problem

The expenditure of households is constrained by their incomes, which are obtained by selling the services of factors of production. But there is a curious difficulty here. In competitive conditions firms will not buy more of a factor than the last or marginal unit is worth to them (in terms of its productive contribution); nor on the other hand, being profit maximizers, will they fail to buy additional units if the productive contribution is worth more than the factor costs. Hence the value of the marginal product must equal the competitive factor price. This applies to all factors, all being governed by the laws of supply and demand, which operate the same way in factor markets as in markets for final products. (Of course the particular elasticities of supply and demand may vary widely.) However there is an important constraint here: The income paid out consists of claims in real terms, and these claims must exactly add up to the real product produced. Suppose the claims issued as factor payments were less than the product. Then the firm would necessarily be the owner of the residual, and could not be in equilibrium. Moreover, part of the product would then be distributed in accordance with a principle other than marginal productivity. Suppose the claims added up to
Introduction. Cracks in the neoclassical mirror: on the break-up of a vision

more than the product. Then the claims – the income payments – cannot be made in real terms, even though each factor receives its real marginal product. At least one factor must end up with less than its marginal product. But which one and why?

We must explore this more carefully. For neoclassical competitive equilibrium to be possible two conditions must be met: factors must receive their marginal products in real terms, and these payments must exactly add up to the total product. These are necessary but of course not sufficient conditions. Both could be met, while factor supplies were out of equilibrium. But if they are not met, a stronger statement can be made: the system should not merely be described as being in disequilibrium. Rather it is a case where the model must be considered an improper representation of neoclassical ideas.

The distinction is subtle but important. A disequilibrium position is one from which, given suitable incentives, agents could, conceivably move towards the equilibrium. By contrast, an unstable position is one in which incentives are so structured that agents’ choices will tend to move them away from equilibrium. A system may possess no equilibrium, meaning that no set of choices can simultaneously satisfy all agents. These are all to be distinguished from the case where the model improperly represents the underlying ideas. In this case the problem arises from an inconsistency between the functional relations between variables postulated in the model, and the meanings assigned to the agents and the variables of the model. When competitive conditions prevail, so that factors receive their real marginal product, and yet the income paid out does not add up to the output produced, the resulting situation is not a disequilibrium, nor has the situation arisen because agents’ motivations are mutually incompatible. The problem is that the output produced in a system of private property always and necessarily belongs to someone, so the assignment of the rights to it and claims over it, must be exact at all times, whether or not that assignment is compatible with the agents’ plans, incentives and motivations. But if the system is supposed to determine earnings on market principles, that is, by supply and demand, then there can be no portion of output falling to a residual claimant, nor can real claims be distributed in excess of output, for that would imply a net liability on the part of the residual claimant, i.e., the firm. Income as a residual contradicts the basic idea that income results from the sale of factor services, that is, that it is a reward for a productive contribution. Note that residual income is not the same as disequilibrium factor reward. Out of equilibrium, we would expect to find factor services being sold in disequilibrium amounts at prices which failed to clear the markets. We would expect disequilibrium incomes to result, that is, incomes greater or less than the factors’ marginal productivity. But that is no reason to expect to find a different kind of income appearing: positive or negative incomes which are not the result (either implicitly or explicitly) of a sale of factor services. Where could such incomes come from? Income in excess of costs will be a net surplus; are there surplus goods as well? Will such income be saved or spent, and what effects will result? Such questions are clearly disruptive; they portend a shift to a different framework of ideas, one which sees production, organized along class lines, characteristically resulting in a surplus appropriated by the dominant class, through the exchange and property system. The magic of neoclassicism lies precisely in its ability to make the surplus disappear; all income results from a sale: it is payment for productive services.

So the problem stems from the need to reconcile two ideas basic to neoclassicism, the concept of the firm as one of the basic agents in the market, and the doctrine that income payments represent a market sale of factor services. The firm as an agent supplying products in the market entails that the firm is the residual claimant – the owner of the total product, including all liabilities incurred in production. But the conception of income as the sale of factor services requires that any residual net claims to product, positive or negative, incurred by the firm, be represented as the proceeds of a sale of a factor service. Generally this has been held to be the service of entrepreneurship, which was, however, never specified in any concrete way. It was only, and could be only, the service of incurring the residual claims. This was seen to involve risk, so the service was “risk-bearing.” But this is simply the service of being the residual claimant, that is, of being the firm.

We can begin to see how interwoven neoclassical theory is. Its parts are interdependent; one doctrine or model cannot be revised, “other things being held constant” for the interdependence is logical. The doctrine that incomes are rewards for productive service, the theory of the firm and the marginal productivity theory of distribution all have implications for one another. For example, the employer, conceived as the risk-bearing residual claimant, is clearly not supplying a factor of production. Hence any earnings received would not represent revenue from a sale of factor services. Wicksell and others therefore argued that the entrepreneurs’ earnings had to be zero, in equilibrium. An alter-
Edward J. Nell

native was to treat entrepreneurship as a factor entering production, which would be supplied as a function of earnings, and whose marginal product would decline.

But then Euler’s theorem would apply; if, when all factors, including entrepreneurship, were increased, returns increased less than proportionately, there would be residual profits, while a more than proportional increase entails residual losses. Who would bear these gains and losses? The entrepreneurs were supposed to be the residual claimants. If entrepreneurship is now a factor entering into actual production, another residual claimant will have to be found who buys the service of entrepreneurship, and takes the ultimate gains or losses. Alternatively we can assume that the production function, including entrepreneurship, always shows constant returns to scale, in which case there are no ultimate gains or losses, and all factors always receive their marginal products.

Consider the first case, where increases in all factors lead first to increasing then to diminishing returns for the firm. The only point consistent with the proposition that income is the reward for the productive services of factors is the minimum point of the average cost curve. This will be the competitive price, and the corresponding quantity will be the competitive output, both determined without reference to demand. Demand for the commodity as a whole will have a role to play in setting the size of the market only if all firms are identical. In that case the long-range marginal cost for the industry will reflect the entry or exit of firms producing at their minimum cost point. As such firms enter producing at their least-cost size, the price will fall, since more is thrown on the market. So long as price is above everyone’s minimum average cost, entry will continue, until price is reduced to average equals marginal cost. Both the size of the market and the number of firms will be determined by demand and supply, though the equilibrium size of each firm is wholly independent of demand. But if there are economies of scale, external to firms but internal to the industry, then as firms enter, not only do prices fall, but so does every one’s average costs. If average costs fall faster than demand price, then no equilibrium may exist. But even when one does there is still a problem in the long run. The long-run average cost curve is the envelope of the short-run curves (which in this case are the curves for the firms); hence when returns to scale for the industry are increasing, the long-run curve will always touch the short-run curves to the left of the lowest point. So, as firms set price equal to short-run marginal cost, which is also equal to minimum short-run average cost, there will always exist long-run profits, accruing as residual windfalls to the firms. But if there are initially economies of scale, external to the firm but internal to the industry, and then at a larger size, similar diseconomies of scale, the size of the industry will also be determined by the requirements of distribution, independent of demand. In either case, price is determined independent of demand, by the requirements of distribution.

Suppose, however, that competition is imperfect. Firms will hire factors up to the point where the marginal revenue product equals the marginal cost of the factor, in money, to the firm. Assume physical constant returns to scale, with perfect factor and imperfect product markets. Then there will be diminishing returns in value terms, since marginal revenue falls. (If the factor market is imperfect, but marginal factor cost rises more slowly than marginal revenue falls, this will remain true.) As a consequence there will be positive profits accruing to the firm as the residual claimant. These profits arise from the combination of constant costs and falling average revenue, and are in no sense a reward for factor services.

Finally, these difficulties reappear at the aggregate level of analysis. Marginal productivitiy theory began as a theory of the demand for factor services, at the level of the firm. But obviously factor services are sold economy-wide, and first-order efficiency conditions require that the marginal value products in all lines be equated, while the second-order conditions require for stability that these marginal value products be diminishing. Observation seems to indicate the presence of conditions broadly approximating those of competition: wage rates between large groups of similar laborers are both equalized and stable. Hence the inference can be drawn that the wage equals the marginal product of labor in aggregate output, and further, that this marginal product diminishes. A similar inference can be drawn for capital, and together with the doctrine that all income represents payments for factor services, this permits the postulate of an aggregate production function, from which the income shares of the factors can be derived. The same kind of difficulties, of course, arise. If the function is not linear and homogeneous, residual income, positive or negative, will exist. But who will get this residual? For the function deals with aggregate capital, and the aggregate capital includes the ultimate ownership of all the firms. Hence any residual is already assigned to capital. In the case of the aggregate production function, the difficulty recognizing the concepts of the firm and factors of production becomes a problem within the concept
of capital itself, since capital is at once a factor of production and represents the ownership of all the firms in the economy.

Perhaps this discussion could be summed up by looking at it formally. The neoclassical approach requires that factor markets be analogous to product markets. Income results from the sale of factor services, and the markets are governed by supply and demand. Looking at the aggregate picture, we have five variables to determine: the real wage; the rate of interest; the amount of labor employed; the amount of capital used; and the total output. But there are six equations to determine them: demand for labor; demand for capital; supply of labor; supply of capital; the production function; and the distributional identity. It is this overdetermination which gives rise to the overpayment or underpayment of income, the positive or negative residual incomes. To eliminate this difficulty, neoclassical theory has adopted assumptions that make the distributional identity depend on the production function, reducing the number of independent equations to five. But this simply creates problems for the theory of the firm since the cost curves of suppliers in product markets reflect the same data that determine distribution.

The firm and factor rewards

How then is the concept of the firm to be reconciled with the notion of income as a reward for the productive contributions of factors? If supply curves are marginal cost curves, and if cost curves are derived from the firm’s production function, then for the long-run supply curve to rise, returns to scale must diminish. This is consistent with Marshall’s dictum that supply and demand are like two scissor blades – both equally involved in the cutting – and it permits the study of both equilibrium for the firm and for the market. But as we have just seen, diminishing returns to scale does not permit a coherent theory of distribution; less than the value of output will be distributed, the remainder remaining in the hands of the firm, leaving the account of the firm, as well as the theory of distribution, in an unsatisfactory state.

The simplest way around the problem has been to adopt Wicksteed’s solution (Wicksteed, 1894), to assume that production functions were linear and homogeneous. This solves the problem of distribution without determining the size and equilibrium of both the firm and the market independently of demand, which is the unfortunate consequence of following the route suggested by Walras (1954) and Wicksell (1934).

And it eliminates the necessity of explaining the existence of residual incomes, not paid for productive services. Moreover, only a linear and homogeneous production function makes sense in the aggregate case, where there cannot be a residual claimant, since aggregate capital includes the equity, and so the ownership, of all firms.

But there is, nevertheless, a heavy price to pay for this solution. Three points spring to mind. First, the assumption of constant returns requires that technical progress be treated as a shift in the production function. But for long-run problems this seems artificial. On what grounds could one distinguish between a shift and a movement along a function? Second, the scissors no longer cut with both blades. The long-run marginal cost curve is horizontal, so prices are fixed independent of demand. Third, under perfect competition the size of the firm and the distribution of output between firms is altogether indeterminate. All three of these, and particularly the third, cut deep into the standard version’s ability to deal with the analysis of particular markets.

If making sense of marginal productivity theory requires one to assume that production functions are linear and homogeneous, there may be little gained by remaining within the format of the standard version. For one thing, that format is open to the Cambridge objections. Yet if production functions are linear and homogeneous, the principal strength of the standard version, its detailed and precise analysis of markets, is largely lost. One might as well go all the way, and adopt the Walras-Cassel approach, leaving the position of the firm indeterminate, but gaining the advantages of mathematical power, coupled with a detailed representation of production.

Costs of abandoning the standard version

Let us examine this more closely. The Walras-Cassel approach is concerned with processes, irrespective of how the process is owned or operated. Yet it is the latter – the conditions the firm faces – which provides the starting point for the theory of market behavior. Most of the main propositions of orthodox economics, at least those normally put forward in textbooks and policy discussions, revolve around market analyses, and often depend on factor market analysis, where the factor is understood in an aggregate sense, for example, labor in general, or capital in general. The standard version is admirably suited for this sort of analytical task, but the activity analysis model is not. There is a cost to orthodox economics in abandoning the stan-
Edward J. Nell

standard version in favor of activity analysis. To see just what this cost is, we must look at the doctrines and disputes of orthodox economics. Simply to illustrate this point, consider three groups of topics, by no means of equivalent theoretical significance, but on which a great deal has been written by leading economists in the last several decades. In the first group let us put those discussions that do not and cannot make sense apart from the conception of income as the reward to a factor’s productive contribution, where each factor is defined broadly as consisting of all those elements that can substitute for one another in a certain production role. The second group comprises discussions that have made use of the broad concept of a factor, and of marginal productivity theory, but that could conceivably be restated on some other basis. The third group concerns the fundamental neoclassical concept of market efficiency and can be stated without reference to the broad conception of a factor. To abandon the standard version, then, is to dismiss the first group altogether, requires rethinking the second, and a shift in emphasis in the third. Let us consider each group of topics in turn.

First, those claims and results that could not conceivably be stated or established without drawing on aggregate marginal productivity theory — that is, that involve economy-wide factor markets, understood and analyzed according to neoclassical principles. Some examples include: the studies of the sources of growth; evaluating the respective contributions of capital, education, and technical progress; the very statement of the problem under study assumes marginal productivity theory. Similarly, the theory of human capital only makes sense if human productive capacities of different kinds can be measured and compared in terms of earning power. The assumption is that earning power measures productive contribution at the margin — not indeed for every individual, but for the average of groups of similar workers. On the basis of this assumption, it can be seen that investments in human capital can be calculated and used to help explain growth, the distribution of earned income, investment in education, and related matters (Thrust, 1970). But if earning power does not measure productive contribution, then the conception simply breaks down. Acquiring a skill may increase earning power, but it cannot be called human capital accumulation unless it enhances productivity. Two other related fields are the neoclassical theory of growth, both one- and two-sector models, an adaptation of which provides the basic framework for the modern approach to international trade — the Heckscher–Ohlin factor price equalization theory. Again it is difficult even to describe either neoclassical growth or trade theory apart from the concept of a broadly based, or aggregate factor of production, the return to the ownership of which measures the factor’s productive contribution. Activity analysis retains the essential idea of a factor of production, but restricts the concept to limited physically identical commodity groups, making it impossible to examine the contributions of labor-in-general or capital-in-general, in situations where prices are variable.

To this first group we should add the studies of portfolio management, in particular the debate over the Modigliani–Miller proposition that the cost of capital to the firm is independent of the firm’s debt-equity ratio, for this proposition is formulated in terms of homogeneous fund capital, and the interest rate, neither of which have a place in activity analysis.

It appears that this first group comprises matters, the discussion of which simply has to be dropped, if neoclassical theory abandons the standard version as unjustifiable. It is not so much that anything in particular is wrong; it is that the problems to be investigated have been defined in terms which have been admitted to lack any sense. If the aggregate production function is an unjustifiable construct, then the marginal products of labor and capital in general have no well-defined meaning, whether we mean the societywide marginal products or those for a particular industry, or specific firm. The point is that when the price changes and the composition and substitution effects entailed by competition have been taken into account, the resulting change in factor earnings bears no orderly relationship to any reasonable and unambiguous measure of productive contribution.

Let us look at the second group of discussions. These certainly depend, in common discourse, on the neoclassical theory of factor pricing, but they seem to depend less strongly on it. They might perhaps be analyzed on the basis of some modified version of the activity analysis model, though no such presentation has so far attained any wide currency. The debate over the Phillips curve is an example. In many, perhaps most discussions, the contention that a change in the level of unemployment will affect the rate of inflation of prices, as opposed to wages (where the effect may be supposed to be direct), depends on the theorem that the real wage is fixed at the level of the full capacity marginal product, so that, assuming competition, price equals marginal cost equals the money wage divided by the marginal product, hence a rise in
Introduction. Cracks in the neoclassical mirror: on the break-up of a vision

wages – productivity constant – raises marginal cost. This presumably justifies the contention that the rate of price inflation equals the rate of wage inflation minus the rate of productivity growth. Critics of the Phillips curve, however, rely no less heavily on marginal productivity theory. The argument that there exists a natural rate of unemployment is usually presented as a denial of workers’ money illusions. An increase in exogenous demand initially raises prices, temporarily lowering real wages, thereby inducing employers to expand employment and production. But workers will revise their price expectations, and incorporate their new perception into their wage demands. Wage rate increases will eventually catch up to prices, raising real wages once again, so that employers will cut back to their original level of output and employment. The original level of unemployment will be restored, but the rate of inflation will be permanently higher.

Both sides of the dispute, those who defend the negatively sloped Phillips curve, and those who maintain that it is a vertical line cutting the horizontal axis at the natural rate of unemployment, state their position in terms of the theory that factor incomes arise from their productive contributions at the margin. Labor’s real wage cannot be permanently raised above or lowered below its marginal product.

The same sort of argument accompanies the discussions of the Pigou effect, where the emergence of unemployment in competitive conditions will lead both wages and prices to fall together, raising the real value of cash balances, so leading to increased spending. Again both wages and prices fall together because in competitive conditions the real wage cannot (except temporarily) deviate from the marginal product. In general, textbook discussions of the Keynesian system present the labor market in terms of a demand curve based on labor’s marginal product.

But it is not clear in any of these cases how much depends on factor incomes being rewards for productive contributions. All that is necessary is that for some reason the real wage must be fixed, so that price and wage inflation have to move together. Marginal productivity theory provides a reason, but if that is disallowed, perhaps another theory could be found. Unlike the first group, where abandoning economy-wide factor markets based on marginal productivity undermined the rationale of the question at issue, in the second group of discussions, an alternative to marginal productivity could perhaps be found, as the “post-Keynesians,” for example, have proposed. Indeed, it is not even necessary to maintain that all income constitutes a reward for productive contribution. So long as the real wage is fixed, while employment and demand vary, profit income could be residual. (By contrast, if profits are a residual, they do not measure the productive contribution of a factor; if a person’s salary reflects the status of his family or school, then it does not measure the productive capacity embodied in his human capital.) However, it might not be easy to adapt the activity analysis model to this case, for these discussions do concern the economy as a whole, hence require an analysis of labor on an economy-wide basis, just as the standard version provides. But activity analysis treats each job separately just as it treats each machine separately. It is difficult to see how this more detailed approach could help.

The third group consists of one proposition, expressed in many different ways. It is the claim that the price system brings about an efficient allocation of resources among competing ends. The efficiency of competitive markets is the centerpiece of neoclassical theory, and the activity analysis approach demonstrates this rigorously, and in vivid detail. Indeed the standard version relying on aggregates in the factor markets cannot do it well; activity analysis can exhibit the productive contributions at the margin of any and every kind of equipment or labor. However, for that reason precisely, activity analysis cannot represent the formation, through competition, of uniform factor prices – a uniform wage for labor of a common level of skill in a wide variety of occupations and industries, or a uniform rate of profit on capital. The standard version is superior, therefore, in this important respect – it represents competition in the factor market better.

Capital in the Walras-Cassel model

This situation has led some economists to view the two approaches as complementary. For some purposes one uses the standard version, for others activity analysis. This is acceptable when the problem concerns a particular event on some definite occasion in a particular market. It is another matter when the issue concerns the nature of the whole economic system. Both approaches offer a general characterization of the competitive market system, and if they are both accepted as revealing different aspects of the truth, then they must be mutually compatible.

The standard version deals with factors in competitive conditions; laborors compete with one another for jobs, and capital funds flow to
where they will receive the highest rate of return. The activity analysis model determines the earnings, for example, of particular machines, or workers. But these earnings are rents, and are subject to requirement that.

[The] permanent fund of capital . . . is put into such forms that the rent secured by one concrete form or capital–good, is as large a fraction of its value as is that received by another . . . This equalizing force determines the number of capital–goods of each kind; and this, again governs the rents they severally earn. (Clark, 1893, p. 125)

The question confronting us is whether the modern activity analysis model is compatible with the capital theory of the standard version.

This can be answered by examining a simple Walras-Cassel general equilibrium model. Let there be a vector of resources, \( r = r_1 \ldots, r_m \), representing nonproduced means of production. Let \( x = x_1, \ldots, x_n \) represent the output of the \( n \) various goods. The input matrix, \( A \) is \( m \times n \), shows the amount of resources per unit output of the goods. The price vector is \( p \) and the vector of rental values of the resources is \( v \). The model can be formulated in programming fashion as follows:

\[
\begin{align*}
\text{max} & \quad p'y \\
\text{subject to} & \quad Ax \leq r \\
& \quad x = F(p, v)
\end{align*}
\]

\[
\begin{align*}
\text{min} & \quad r'v \\
\text{subject to} & \quad A'v \geq p \\
& \quad v \geq 0
\end{align*}
\]

The function \( F \) gives the demands for the goods, \( x \), as a function of prices and the incomes of households received from the rental of their resources.

A principal concern of those who have worked with this model has been to show that it possesses a solution – that a general equilibrium will exist under plausible conditions.\(^{14}\) Let us assume that such an equilibrium exists, and interpret the model further. A crucial question concerns the meaning of the resource–vector, \( r \). These resources enter into production but are not themselves produced, and evidently need no renewal. For if they were used up or worn out, the equilibrium determined would simply be a one-shot affair, or not even that. To be an equilibrium, an economic configuration must not only be capable of persisting, it must also account for all the market-related costs and benefits. But if using the sources uses them up or wears them down, then such user cost must be represented in the model.

Are there any resources capable of being appropriated and marketed which enter into production, but are not used up?\(^{15}\) Ricardo spoke of “the original and indestructible powers of the soil,” but after the dust bowls of the 1930s we should know better (Ricardo, 1973). Space or location might be an example, but such things enter production more in a metaphorical than an engineering sense. In general, resources entering production are used up or worn out, and will eventually need replacement. Unless this is shown, the model simply lacks application, not because it is too abstract, or unrealistic, but because it misrepresents the nature of production. It is guilty of an error of commission, not merely, as the phrase unrealistic suggests, of omission.

Yet this is easily remedied. Let some of the commodities produced be the same goods as the resources. Resource holdings can represent the initial endowments, and the coefficients then represent the user cost, which must be made good by replacement.

The implications of admitting that some of the goods in the initial endowments may be produced means of production are dramatic. Consider a simple model, containing only corn and iron:

\[
\begin{align*}
\text{max} & \quad x_2p_2 + xyp_1 \\
\text{subject to} & \quad a_2x_2 + ayp_2 \leq C \\
& \quad a_2x_2 + ayp_2 \leq I \\
& \quad x_2, x_1 > 0
\end{align*}
\]

\[
\begin{align*}
\text{min} & \quad Cv + Iv \\
\text{subject to} & \quad a_1x_1 + a_1y_1 \geq p_1 \\
& \quad a_1x_1 + a_1y_1 \geq p_1 \\
& \quad x_1, x_2 > 0
\end{align*}
\]

Let the demand function \( F(p, v) \) be such that prices satisfying it result in the production of both corn and iron, and further suppose that these quantities permit the necessary replacement to take place. No loss of generality is involved; this merely guarantees a solution compatible with replacement. Consider the diagram (Figure 1.1) first of the maximizing, then of the minimizing problem. So long as the slope of the price ratio lies between the slopes of the two constraints, both goods will be produced. If both goods are produced both resources will be assigned positive values. The minimizing problem yields these shadow values, which represent the rentals that can be obtained from the commodities. The slope of the line joining the optimal value of \( v_t \) to the optimal value of \( v_e \) gives the ratio of the rental values. The slope of \( v_t/v_e \) need not equal \( p_1/p_2 \). If it did, \( p_1/p_2 \) could be changed without changing \( v_t/v_e \) so long as the change stays within the limits prescribed by the slopes of the constraints.\(^{16}\) But if \( (v_t/v_e) \approx (p_1/p_2) \), then \( (v_t/v_e) \approx (v_t/v_e) \), that is the ratio of rental earnings to supply price, for the two goods will be different. But capitalists will invest