SOME RECENT DEVELOPMENTS IN THE THEORY OF PRODUCTION

ROBERT M. SOLOW
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ONLY two or three years ago, Walters published a long survey of "Cost and Production Function" (38).[1] Its bibliography listed 345 items, from Alchian to Zellner. It is true that Walters gave much of his space to econometric work (this is a field in which theory and empirical work are, happily, hard to separate). It is true also that one has the casual impression of having seen another 345 items go sailing by since 1963. Nevertheless, I do not intend to attempt another rounded survey of even the pure theory of production. Instead, what follows is merely a brief summary of a handful of new ideas that seem to me especially worth attention. Any such selection is bound to be idiosyncratic. It invites difference of opinion about what is useful or interesting and what is merely neat. That may be inevitable; in any case, these are a few theoretical leads that seem to me to be worth following up, perhaps to establish their full utility, perhaps eventually to dismiss them as dead ends.

Before I begin, I must justify some exclusions. Despite the substantial volume of recent work on the static theory of production, it seems to me that little has occurred to alter the standard textbook accounts. If one were rewriting now the chapter on cost and production in Samuelson's *Foundations of Economic Analysis* or the corresponding parts in the standard texts of Boulding or Stigler, there is not much one would have to change. There has been, in fact, one major advance in the pure theory of production in the past twenty years, and that is the development of the linear model of production under the heading Activity Analysis or Linear Programming. The texts I read as a student were at home only under the assumption that different inputs were smoothly

[1] Figures in parentheses refer to the bibliography at the end of this paper.

substitutable for one another in production. (Samuelson's *Foundations* is only a partial exception.) Today one can handle with ease and elegance the case of a finite number of joint production activities. Indeed, if it comes to that, computational possibilities are better for this case than for the older one.

As I have said, I take the development of linear theory to be a major advance in the theory of production. I think it is fair to say, however, that the results of activity analysis and linear programming have gone to confirm and considerably to deepen the main insights of the smooth theory of production. They have not changed the results in any sharp or unpredictable way. In any case, I shall not discuss activity analysis further. There have already been several full expositions of linear theory in the context of the theory of production: Dorfman, Samuelson, and Solow (9), and Boulding and Spivey (4); and the material is beginning to be included in textbooks, e.g., Davidson, Smith, and Wiley (6), and Dorfman (8).

The pure theory of production is fundamentally microeconomic in character; it deals with physically identifiable inputs and outputs. In the classroom one usually says that the economic theory of production takes for granted the "engineering" relationships between inputs and outputs and goes on from there. By contrast, much (though not quite all) of the recent interest in the theory of production has been macroeconomic in character. Since the "inputs" and "outputs" are statistical aggregates like "labor," "plant," "equipment," "durable manufactures," there is no possibility of finding engineering relationships. Econometric methods have to do duty instead. Still, it remains an intriguing idea to deduce economically useful production functions from raw technological information. I had hoped to say something on that subject; but it appears that very little of general interest has been done since the early paper of Chenery (5). I do not mean to deprecate the extensive work on agricultural input-output relations, summarized in Heady and Dillon (18), or the process analysis of Manne, Markowitz, and others (26), but they seem to me to be mainly interesting as empirical rather than theoretical enterprises.

I had hoped to find an exception in the interesting idea of Kurz and Manne (24). They begin with some "engineering" estimates of the capabilities of 115 different types of machine tools (e.g., "boring machine, horizontal, under 3″ spindle, under 36″ bed") to perform each of 129 different metalworking tasks, defined in terms of geometrical shape, size of piece, tolerance, and size of lot (e.g., "flat surface—no contours, small size, semiprecision, long run"). Obviously, some combinations of machine tool and task will be hopeless, but for the rest Kurz and Manne use process-analysis estimates of output per worker and machine tool in number of pieces per daily eight-hour shift, and capital investment per worker and machine tool. From this raw material they try ingeniously to estimate a production function which will adequately summarize the possibilities of substituting one machine for another on different tasks and the more general possibility of substituting capital investment for labor in metalworking tasks. Unfortunately, the execution seems to be faulty and the criticism of Furubotn (14) justified. To get on with the job, Kurz and Manne eliminate as "inefficient" all those task-machine combinations that require higher investment costs per machine tool and worker without yielding higher output per machine tool and worker. But for this sort of step to make sense, capital costs have to be "annualized" or, in this case, "diurnalized"; depreciation rates and maintenance costs need not be equal or proportional for machine tools with different initial costs. Even if that were accomplished, it would remain true that changes in the relative prices of alternative machine tools would change the shape of the Kurz-Manne "production function" and even the composition of the set of efficient machine-task combinations. This means that the Kurz-Manne production function is less like an "engineering" relationship and more like a macroeconomic relation than might casually appear. I do not find this conclusion so very disturbing. The problems are not different in principle from those arising in any attempt to construct a conglomerate measure of input or output; and perhaps machine-tool prices move more or less together in fact. If aggregation is inevitable, relax and enjoy it.

It will come as no surprise to any reader of the current literature that many of the topics I do intend to mention have to do with the representation and analysis of technological change. I propose to begin with the idea of describing shifts in the production function as "factor-augmenting," and go on, as an application, to summarize briefly some first steps toward a theory of induced bias in technical progress. I then turn to Kaldor's notion of replacing the production function altogether by a "technical progress function" and to Arrow's extension of the production function to include "learning by doing." Then I want to call attention to some recent work on the definition and significance of the elasticity of substitution when there are more than two inputs, to the construction of models of production in which the choice of input intensities is once and for all with no further variation possible, and finally to a neglected paper of Houthakker's (19) on aggregation.

The microeconomic theory of production merges imperceptibly into the theory of distribution; it is hard to write about one without trespassing on the other. Analogously, when the theory of production is cast in macroeconomic terms it merges imperceptibly with the theory of economic growth. I shall try hard to avoid poaching, if only because this territory has recently been so superbly surveyed by Hahn and Matthews (17). To stay clear of growth theory I must limit myself so far as possible to the descriptive aspects of the theory of production, and I must avoid discussing equilibrium conditions to the extent that is possible.

Factor-augmenting Technical Change

Leave aside for the moment all fancy considerations about the possible "embodiment" of technological change in capital goods, trained labor, or anything else. Then a general way to represent technological change in a single-product production function is:

$$Q = F(X, Y; T) \tag{1}$$

where Q is output, X and Y are inputs (all measured in natural physical units), and T is a parameter or even, for extra generality, a vector of parameters, each value of which corresponds to a different level of technology. It is natural to think of T as changing in time, perhaps smoothly, as knowledge accumulates; in that case F should be a nondecreasing function of T. But there is no reason why T should not change by discrete jumps, or from place to place, or climate to climate, or from entrepreneur to entrepreneur. Then there is no need for F to be monotone in T. In principle, the production functions corresponding to two different T's can be *any* two production functions. If we assume constant returns to scale in X and Y, so that everything is summed up in the unit isoquant, then corresponding to T_1 and T_2 may be any pair of isoquants for $Q = 1$. If the change from T_1 to T_2 is intended to be unambiguous technical progress, then the only restriction is that the later isoquant should never pass outside of the earlier. If in fact the shift from T_1 to T_2 has occurred continuously, then any continuous deformation of the first isoquant to the second is a possible path, provided only that the movement is always inward.

This is rather too general for use, and some specialization is in order. It seems to be helpful in theory and in empirical practice to represent technical change as factor augmenting:

$$Q = F[a(T)X, b(T)Y] \tag{2}$$

It must be realized that this is a genuine specialization; not every (1) can be written as (2). The loss of generality is indicated by the fact that (1) makes output a function of three variables, while (2) is a function of two, only the variables themselves are $a(T)X$ and $b(T)Y$, which bear a natural interpretation as inputs of X and Y in "efficiency units." Obviously if a increases by 5 per cent and b by 10 per cent, then one unit of X and one of Y can do exactly what 1.05 units of X and 1.10 units of Y could do before the change. It is tempting to think that because a change in a is X-augmenting it must be, so to speak, X-specific; for example, if X is homogeneous labor one might expect an improvement in its quality to be reflected in an increase in a. But this is an error. An improvement highly specific to X may be reflected in a or b or both.

It is well known that (2) is itself a generalization of the definitions of "neutral"[2] technological change proposed by Hicks and Harrod. Hicks's definition was that the marginal rate of substitution between X and Y should be independent of T for each fixed X and Y; Harrod's was that the average product of X should be independent of T for fixed marginal product of X. It has been often shown that Hicks-neutrality is equivalent to (2) and $a(T)/b(T)$ = constant, while Harrod-neutrality is equivalent to (2) and $a(T)$ = constant. (All this is under constant returns to scale.) One asks immediately: On what grounds is the requirement $a(T)$ = constant to be preferred to the perfectly symmetrical alternative $b(T)$ = constant? The answer is that the Harrod definition is framed particularly with the idea in mind that X is a produced factor of production—capital—and Y is a primary factor of production—labor. Then Harrod-neutral, or purely labor-augmenting, technical progress is an especially convenient vehicle for the study of steady economic growth, though there is room for argument about its factual plausibility. The symmetric assumption of purely capital-augmenting technical progress has certain nice properties of its own, mainly in connection with aggregation [see Fisher (13), Gorman (15), and Samuelson (33)].

One of the advantages of formulation (2) is that it gets away from these unnecessarily tight restrictions to "neutrality" of one kind or another, without going all the way back to (1). Diamond (7), Fei and Ranis (11), and no doubt others have proposed to describe the course of technical progress by an index, say R, of the rate of progress, and an index, say B, of its bias. As applied to the general formulation (1), Diamond suggests

$$R = \partial \log F/\partial T$$

$$B = \frac{\partial}{\partial T}\log\left(\frac{\partial F/\partial X}{\partial F/\partial Y}\right) = \frac{\partial}{\partial T}\left(\frac{F_x}{F_y}\right)\Big/\left(\frac{F_x}{F_y}\right)$$

This particular index of bias is obviously Hicks-oriented. It shows what happens to the marginal rate of substitution between X and Y for fixed X and Y as the level of technology T changes; Hicks-neutrality means $B = O$. One could easily define a Harrod-oriented index of bias, as I shall show in a moment. Calculation shows that

$$B = \left[\frac{a'(T)}{a(T)} - \frac{b'(T)}{b(T)}\right]\left(1 - \frac{1}{\sigma}\right)$$

$$R = \eta_X \frac{a'(T)}{a(T)} + \eta_Y \frac{b'(T)}{b(T)}$$

where σ is the elasticity of substitution between X and Y, and η_X and $\eta_Y = 1 - \eta_X$ are the elasticities of output with respect to X and Y respectively. (Observe that $B = O$ when either a/b is constant, even momentarily, or $\sigma = 1$, in which case a and b cannot be distinguished.) I leave it to the reader to reason out why the direction of bias depends on whether σ is greater or less than one.

Fei and Ranis (11) and Sheshinski (34) have introduced a Harrod-oriented index of bias through the formula

$$C = \frac{\partial}{\partial T}\log\frac{Q}{X}$$

evaluated with F_X constant. The same sort of calculation for the factor-augmenting formulation yields

$$C = (1 - \sigma)\frac{a'(T)}{a(T)},$$

so that $C = O$ if and only if a is constant or $\sigma = 1$. By symmetry there is an analogous measure of bias which vanishes when b is constant:

$$D = (1 - \sigma)\frac{b'(T)}{b(T)}$$

There is, of course, a lot of redundancy here. As Diamond, Fei and Ranis, and Sheshinski have shown, the conventional analysis of the production side of economic growth can be carried on in terms of η_X, σ, and any two of R, B, C, D. For example,

$$D - C = \sigma B$$

and

$$\eta_X C + \eta_Y D = (1 - \sigma)R.$$

It should be kept in mind that neither η_X nor σ is a constant, independent of X and Y, except in special cases.

I want to emphasize that the value of the factor-augmenting representation is not in such taxonomic identities as these. It is that in empirical as in theoretical work it gives "something" to estimate or talk about, namely, the functions $a(T)$ and $b(T)$. It does this, as I have mentioned, only at the cost of some generality; the factor-augmenting representation is not broad enough to encompass changes in the elasticity of substitution or in Cobb-Douglas exponents, for example. On the other hand it does free the discussion from the straitjacket of "neutrality" of one kind or another; this is an advantage if, as seems to be the case, neutrality is too restrictive to fit the facts.

The extra flexibility is especially valuable if one has to account for more than two factors of production. Of course, the whole analysis becomes more complicated; one needs an index of the rate of technological progress plus two indexes of bias plus three elasticities of substitution, and there is even some choice about how to define the elasticity of substitution. Besides, if the context is economic growth, the only possible extension of the underlying idea of Harrod-neutrality turns out to be extremely limiting. One must suppose that there is only one primary factor, all the rest being themselves produced. If one requires that the average products of the produced factors all be constant and independent of the level of technology when all their marginal products are constant, then technological progress must augment only the single primary factor. This seems rather too special. Hicks-neutrality extends more easily, but still seems special. The factor-augmenting representation $Q = F[a(T)X, b(T)Y, c(T)Z, \ldots]$ generalizes easily and may yet be usable.

Induced Bias in Technical Progress

As an example of the theoretical convenience of the factor-augmenting assumption, I shall cite some very recent progress with an old and worrisome problem: the notion of induced bias in invention. The main papers are by Kennedy (23), Samuelson (32), and Drandakis and Phelps (10); Samuelson gives references to earlier literature, particularly Fellner (12). Before Kennedy, the discussion suffered from lack of an explicit representation of the set of inventions or lines of invention among which inventors choose (or at least along which they choose to search). The factor-augmentation functions $a(T)$ and $b(T)$ provide sufficiently—but, one hopes, not laughably—concrete "objects" for the theory to be about.

The basic device of the newer theory is an "invention possibility frontier" which can be written

$$I(a'/a, b'/b) = 0$$

Technical change is still taken to be autonomous, in the sense that there is no accounting for the resources used up in research. There is, however, an opportunity cost, in the sense that only a limited improvement in technology is possible per unit of time, and only one "direction" of improvement can be pursued at a time. Directions of technological progress are described by the factor-augmentation functions, which are now taken as functions of time rather than of some latent level of technology. It is natural to assume that the invention possibility frontier describes a curve in the plane of $a'(t)/a(t) = g_a$ and $b'(t)/b(t) = g_b$ which is falling and concave, like any transformation curve. Drandakis and Phelps insist that we consider only the quadrant where g_a and g_b are both nonnegative. Their reason is that if either is negative there will be some factor proportions for which the new production function is "worse" than the old. I am inclined to think this is unwise. For a particular pair g_a and g_b to represent unambiguous technical progress at every factor ratio, what is required is that the quantity R defined above be nowhere negative. This condition may be satisfied even with one of g_a or g_b negative, if the relative shares happen not to run the full gamut from zero to one; Cobb-Douglas is an extreme example. If all relative shares are possible, then of course Drandakis and Phelps are right. Even so, it is probably better then to step a bit outside the pure factor-augmentation assumption and allow either g_a or g_b to be negative at least where it does not imply technological regress. The reason is that empirical work might conceivably throw up indications of negative factor augmentation; rationally aimed research might do this on the sensible presumption that it need only worry about relative shares not too far from the current ones.

The new theories of induced technical progress operate on the assumption that the economy "chooses" among the combinations satisfying $I(g_a, g_b) = 0$ a best pair. "Best" is usually defined to mean maximizing the instantaneous rate of technical progress R at the going factor shares; if factor-price imputation is competitive, this is the same thing as maximizing the instantaneous rate of decrease of unit costs at going factor prices. (Fellner and Samuelson discuss some longer-sighted criteria.) Maximization of $\eta_X g_a + \eta_Y g_b$ subject to $I(g_a, g_b) = 0$ yields the necessary condition

$$\frac{\eta_X}{\eta_Y} = \frac{\eta_X}{1 - \eta_X} = \frac{I_a}{I_b},$$

where I_a and I_b are partial derivatives. It is to be remembered that η_X and η_Y are, under constant returns to scale, functions of aX/bY. Thus, given the current values of a, b, X, and Y, the rates of growth of a and b are determined.

This is only half the story. Add a mechanism governing the evolution of X and Y (if X is capital and Y labor this amounts to a determinate theory of investment and an assumption about the growth of the labor force) and the story is complete. At the next instant we have determinate values for a, b, X, and Y and therefore enough information to determine the new g_a and g_b, and carry the story further forward. To tell in detail how it comes out would carry me across that narrow line into the theory of growth and distribution. It is useful, however, to say this much. If the elasticity of substitution between X and Y (which is the same as the elasticity of substitution between aX and bY) is less than one, the dynamic process usually tends to a stable limit in which g_a and g_b are constant and therefore relative shares are constant. The reason is that when $\sigma < 1$, an increase in aX/bY decreases η_X; given the concavity of the invention possibility frontier, a decrease in η_X decreases g_a and increases g_b. Suppose, to take an uncomplicated case, that X and Y have different but exogenous rates of growth. If aX/bY is initially growing, η_X will fall, so will g_a, and this process must continue until aX/bY is constant. (If X is capital and Y is labor, this seems to establish a presumption in favor of labor augmentation.) Clearly, if $\sigma > 1$ everywhere this steady state is unstable, and the system will diverge to an extreme of distribution and/or biased technical change.

This particular episode in the theory of production and technical progress has only just begun; there is plenty of room for refinement and improvement. For the fairly short run one probably wants some version of "embodiment." For the long run one can hardly take the invention possibility frontier as stationary; it can itself be shifted by devoting resources to research, and one expects the internal logic of science itself to create a variable "natural" drift of technical change in one direction or another. This might make the invention possibility frontier a function of calendar time, or perhaps of its own past. Samuelson has pointed to a deeper problem [see also Salter (31)]: factor proportions themselves may enter the invention possibility function, if only through the relative shares. The idea is that if labor represents 70–80 per cent of total costs, it offers a larger target to shoot at than capital or other factors. In one version at least, as Samuelson shows, this formulation eliminates the presumption in favor of labor augmentation mentioned above.[2] It is depressing to think how hard it will be to get any empirical light at all on these questions.

It is easy formally to extend this theory of induced invention to three or more factors. But as so often happens it loses transparency. In this case it happens for two reasons: It is hard to capture what is captured by the (single) two-factor elasticity of substitution by any definition of $[n\ (n-1)/2]$ n-factor elasticities of substitution; and the dynamics de-

[2] One symmetrical way to capture this idea is to write (with Samuelson) the invention possibility curve as $I(g_a, g_b, \eta_X) = 0$ with $I_a(g, g, \eta_X)/I_b(g, g, \eta_X)$ identically equal to $\eta_X/(1 - \eta_X)$ where $I(g, g, \eta_X) = 0$. Then the theory yields $g_a = g_b$ all the time, their common value depending on η_X, which itself depends on factor supplies in the conventional way.

pends on a much more complicated interaction of the various substitution possibilities and the various trade-offs along the invention possibility surface.

Kaldor's Technical Progress Function

It is a recurrent theme of modern production theory that technological progress is somehow embodied in or otherwise bound up with investment in capital goods. Incorporation of this hypothesis into a production function like (1) or (2) has led to so-called vintage models of production. These have been so widely analyzed and discussed that I shall not try to survey them here. In an attempt to get at the same phenomenon, Kaldor (21) has proposed a rival formulation. He argues that the connections among production, investment, and technical change cannot be expressed by any kind of reversible relation between inputs and outputs, but can be described by what he calls a technical progress function. This alternative formulation seems to have attracted very little attention and to have inspired no empirical work. I am half inclined to conclude this reflects the Darwinian process at work. On the other hand, the notion of a technical progress function has recently been refined in Kaldor and Mirrlees (22) to the point where it is not really so different from "conventional" formulations based on the vintage model of production. The refinement, however, has not been carried far enough.

In the original version the technical progress function was superimposed on a setting of homogeneous capital and labor. Let me adapt my earlier noncommittal notation by taking X as capital and Y as labor. Let $q = Q/Y$ and $x = X/Y$ and represent relative time rates of growth by $g_q = q'(t)/q(t)$, etc. Kaldor's proposal was to write

$$g_q = K(g_x); \tag{3}$$

he assumed $K(0) \geqq 0$, $K' > 0$, $K'' \leqq 0$. Thus, with capital intensity constant, productivity would increase through general technological drift, but an increase in capital intensity would be associated with a still faster increase in productivity subject to a kind of diminishing returns. I do not find this wholly nonsensical, though it is rather implausible that the relation between the rate of growth of productivity and the rate of growth of capital intensity should necessarily be independent of the degree of capital intensity already achieved.

Many people have remarked that a relation like (3) can be deduced from (1) or (2), but only on the assumption that the production function is Cobb-Douglas. See, for example, Black (3). From (2), for instance, under constant returns to scale, follows

$$g_q = \eta_X g_x + R$$

In general, of course, η_X is a function of x, which does not appear in Kaldor's technical progress function. If (2) is Cobb-Douglas, however, then η_X is constant and we have a technical progress function which is linear, or at least linear at each instant of time; and conversely. But this coincidence seems to me to be uninteresting. It is a factual question how one can legitimately represent production relations when technological knowledge is changing. The interesting analytical questions arise only after we have an acceptable description of technically feasible possibilities, when we introduce a mechanism by which the individual or the economy chooses among them. As Weizsäcker has shown (39), this part of the story is defective both in Kaldor and Kaldor and Mirrlees even if one accepts the technical progress function. Weizsäcker goes on to introduce a criterion of choice and works out a distribution theory on general supply-demand principles. In the same spirit, but without going that far, I think it will be useful to suggest by a simple example what sort of technically feasible choices are implicit in the Kaldor formulation.

To do so, it is simpler to work in discrete time. The analogue of (3) is:

$$\frac{q_t - q_{t-1}}{q_{t-1}} = K\left(\frac{x_t - x_{t-1}}{x_{t-1}}\right)$$

More compactly,

$$q_t/q_{t-1} = 1 + K\left(\frac{x_t}{x_{t-1}} - 1\right) = J(x_t/x_{t-1})$$

Thus, by iteration

$$\begin{aligned} q_t &= q_{t-1}J(x_t/x_{t-1}) = q_{t-2}J(x_{t-1}/x_{t-2})J(x_t/x_{t-1}) = \cdots = \\ &= q_0 \prod_{k=1}^{t} J(x_k/x_{k-1}). \end{aligned}$$

This illustrates that production possibilities at any one point of time depend on the whole path the firm or economy has followed in the past. Arrow's "learning by doing" (1) shares this characteristic; so does the model of Solow *et al.* (36), which has strictly exogenous technical progress.

Imagine a central planning board which has just inherited an economy and its past at time zero; the board looks ahead for, say, two periods, during which time the supply of labor is given. Then production possibilities are

$$Q_1 = Y_1 \frac{Q_0}{Y_0} J\left(\frac{X_1 Y_0}{X_0 Y_1}\right); \; Q_2 = \frac{Y_2}{Y_0} Q_0 J\left(\frac{X_1 Y_0}{X_0 Y_1}\right) J\left(\frac{X_2 Y_1}{X_1 Y_2}\right)$$

In particular, if the supply of labor is constant,

$$Q_1 = Q_0 J(X_1/X_0); \; Q_2 = Q_0 J(X_1/X_0) J(X_2/X_1)$$

Various elementary questions can be answered. For example, suppose there is no depreciation and the planning board has already decided how much it intends to invest over the next two years, so that X_2 is fixed: How should it allocate the total of investment ($= X_2 - X_0$) between the two years to make end-period output Q_2 as large as possible? A necessary condition for a solution with positive investment in both years is

$$(X_1/X_0)\frac{J'(X_1/X_0)}{J(X_1/X_0)} = (X_2/X_1)\frac{J'(X_2/X_1)}{J(X_2/X_1)}$$

This condition is obviously satisfied if the stock of capital is made to grow at the same geometric rate in the two periods; if the *elasticity* of J is monotone the condition is satisfied only then. But at the steady growth solution the sign of the appropriate second derivative is positive if the elasticity of J is increasing with its argument, negative if decreasing. Only in the latter case is steady growth of capital the maximizing strategy. Otherwise (I stick to monotone elasticity to avoid a tiresome catalog) the best strategy is to pile all the investment into one year, no matter which. Black seems to have believed that steady growth was optimal here so long as K was concave, but that is erroneous. It is amusing that if J has a constant elasticity, i.e., $J(u) = ku^j$, then even with varying labor Q_t is proportional to $k^t X_t^j Y_t^{1-j}$, independent of the intervening path. So it is equivalent to the original conventional vintage model. The reader can easily work out the similar case when the planning board wants to maximize $P_1Q_1 + P_2Q_2$ or a similar weighted sum of consumption in the two periods and terminal capital.

The Kaldor-Mirrlees version restricts the technical progress function to a vintage model of production and then asserts that the rate of change of output per man as between last year's equipment and this year's equipment is an increasing concave function of the rate of change of investment per man as between those employed using last year's and this year's equipment. In obvious notation,

$$Q_t = Q_0(L_t/L_0) \prod_{k=1}^{t} J\left(\frac{I_k L_{k-1}}{I_{k-1} L_k}\right)$$

where L_k is to be identified as the labor actually employed operating the kth period's investment. One should ask and answer the same sort of planning questions for this version. I do not take the time to do so because the problem is now a bit more complicated. The difference is that before one could legitimately take total labor as exogenous, but now L_t can be made to exceed the natural increase of the labor force by discretionary scrapping of old, unproductive plant. That there is a family resemblance between the two formulations is revealed by the special case of constant-elasticity J; for then Q_t is proportional to $k^t I_t^j L^{1-j}$ regardless of what has happened since $t = 0$. It is thus equivalent to the standard "putty-clay" vintage model.

Arrow's Learning by Doing

The notion that technical progress could be "embodied" in capital goods was invented to give expression to the common sense picture that many advances in technical knowledge can affect production only when they are designed into new capital goods through gross investment. (I take it as a major intellectual puzzle, by the way, to explain why a notion that seems so self-evident in micro terms should contribute so little additional explanatory power in econometric macromodels. Can we be that close to a steady state?) In such models more investment means higher over-all productivity. But even in such models, the accumulation of technical knowledge is assumed to be autonomous. Many economists have had the idea that technological progress itself has an endogenous aspect, not simply in the sense that society can devote scarce resources to research, but in the somewhat vaguer sense that what happens in production itself has an important effect on the generation of new knowledge about production. Something like learning or exploration may occur. (The development of consumer preferences can be approached the same way, requiring modification of the standard picture of consistent, given tastes defined over the whole commodity space.)

I suppose that Kaldor's technical progress function is an attempt to capture this idea of the endogenous generation of technological knowledge. It seems to me to be defective: Why on earth should the rate of increase of productivity depend only on the rate of increase of investment per man? Arrow (1) has proposed a better thought-out alternative way of capturing much the same notion. His results have been generalized and extended by Levhari (25). Otherwise, apart from mere footnote references, there has been no further development along the lines opened by Arrow. Whether "learning by doing" is a blind alley or merely awaits some concentrated theoretical and empirical effort, I have no way of knowing. It seems at least to be getting at an aspect of reality.

Arrow's particular assumption is that technological change grows out of "experience," and cumulated experience is measured by cumulated gross investment. At any given level of technology there are fixed coefficients in the production of aggregate output from labor and existing capital goods. I adapt my earlier notation so that Y_t now represents cumulated gross investment since the economy began. If $m(Y)$ represents the fixed complement of labor with a unit of capital constructed at a moment when cumulative gross investment is Y (i.e., with serial number Y), and $n(Y)$ is the capacity embodied in a unit of capital of serial number Y, then the fixed coefficient technology implies at each instant of time

$$Q_t = \int_{Y_{t'}}^{Y_t} n(Y)dY$$

$$X_t = \int_{Y_{t'}}^{Y_t} m(Y)dY$$

Where $Y_{t'}$ is the serial number of the oldest capacity actually in use at time t. (The assumptions will be enough to guarantee that the serial numbers of the capacity in use form an interval.) Now if M and N are the indefinite integrals of m and n, one can eliminate $Y_{t'}$ to get

$$Q_t = N(Y_t) - N\{M^{-1}[M(Y_t) - X_t]\}$$

This can serve as a sort of aggregate production function. It is a novel one because its arguments are current labor input and cumulative gross investment, including some capital goods no longer surviving. The last remark is an important one; it is the essence of this model that even the "Titanic" is still contributing to maritime productivity. Even if it can no longer carry passengers, the fact that it was once built makes all current serial numbers a little bigger than they would otherwise be and therefore all current capital more productive than it would have been if the "Titanic" had never existed.

The Arrow "production function" may be a little more transparent in the special case he analyzed, where $n(Y) = n =$ constant and $m(Y) = mY^{-h}$. With fixed coefficients all technical change is factor-augmenting; this is the Harrod-neutral or pure labor-augmenting case, with the difference that the degree of labor augmentation depends on cumulative gross investment. Carrying out the calculation gives

$$Q = nY\left[1 - \left(1 - \frac{1-h}{m}\frac{X}{Y^{1-h}}\right)^{1/1-h}\right], \qquad h \neq 1$$
$$= nY(1 - e^{-X/m}), \qquad h = 1$$

In this form one sees easily what is true in general, that there are increasing returns to scale in the variables X and Y, though the microscopic technology has constant returns to scale—is, in fact, linear. This fact directs attention to what is probably the most interesting consequence of the model for general economic analysis. Under these assumptions about technology, smoothly functioning competitive markets would impute to fully employed labor a wage equal to its social marginal product. The residual quasirents yield a private return to capital which is definitely less than the social rate of return on investment. The builders of the "Titanic" have no way of earning anything corresponding to its posthumous contribution to output (nor to some part of its contribution even before it sank).

Levhari is able to extend the analysis beyond the fixed coefficient case. Let $I(v) = Y'(v)$ be the rate of gross investment at time v, $Q(v, t)$ be the output produced with its use at time t, and $L(v, t)$ be the labor allocated to it at time t. Then Levhari treats the case of an arbitrary constant-returns-to-scale technology with

$$Q(v, t) = F[I(v), Y^n(v)\, L(v, t)]$$

Note that the endogenously generated technical progress is still purely labor-augmenting in this formulation. But factor proportions are now variable, both at the planning stage and after concrete capital already exists. The broad qualitative properties of the model are not much changed, but are somewhat enriched. Levhari treats both the case where $F_2(1, 0)$ is bounded and the case where it is not. In the first case, old capital is eventually retired for economic reasons; in the second case, it is not.

One can think of other generalizations that ought to be carried out. The restrictive assumption of Harrod-neutrality may be necessary if the analysis is to center on steady states, but there is every reason to go further. Advances in technique may be generated as much by employment as by investment; on the other hand one might imagine that the "learning" associated with a given amount of investment might be less if it has to be diffused over a larger number of workers.

The Elasticity of Substitution

In any two-factor production function like (1) or (2) it is handy to have a measure of the ease with which X and Y can be substituted for one another. The standard measure is the elasticity of substitution, defined as

$$(4) \quad \sigma = \frac{F_X F_Y}{F F_{XY}} = \frac{F_X(X, Y, T) F_Y(X, Y, T)}{F(X, Y, T) F_{XY}(X, Y, T)} = \frac{F_X(X/Y, 1, T) F_Y(X/Y, 1, T)}{F(X/Y, 1, T) F_{XY}(X/Y, 1, T)},$$

with the subscripts indicating partial derivatives. (The last step depends on homogeneity or "homotheticity.") It is well known that, thus defined, σ is the (positive) elasticity of X/Y with respect to P_X/P_Y along an isoquant. That is, corresponding to a 1 per cent change in the price ratio or slope or marginal rate of substitution (with output constant) is a σ per cent change in the opposite direction in the ratio of the factors. Thus, for example, if $\sigma = 2$, a fall of 1 per cent in P_X/P_Y is associated with a 2 per cent increase in the ratio of X to Y, and therefore with a 1 per cent increase in $P_X X/P_Y Y$. In other words, the competitively imputed share in output of the more rapidly growing factor rises. Vice versa, if $\sigma < 1$. In general the elasticity of substitution varies from one point on the unit isoquant (and therefore on every isoquant, under constant returns to scale or even slightly more general assumptions) to another. It can oscillate from one side of unity to the other without violating the usual convexity conditions. The cases where the elasticity of substitution is in fact constant all along the isoquant have been much studied for convenience.

Most discussion of the elasticity of substitution has originated in its significance for competitive distribution (because capital grows faster than labor, usually). But there is something to be said for its purely descriptive utility. For example, it will be remembered that the particular parameter σ turned up quite naturally in the earlier discussion of biased technical progress. If competitive imputation were wholly irrelevant, one might still want to have some neat way of describing the degree of complementarity or substitutability between factors, if only for empirical work.

Unfortunately, as soon as one recognizes three or more factors of production it is no longer so clear how one ought to measure the degree of substitutability among them. There are alternative reasonable definitions of "the" elasticity of substitution, each answering a slightly different question. This multiplicity has been known for a long time, but has come to the surface again in the course of the exploration of production functions for which the elasticities of substitution—however defined—are constant. The basic references are Arrow *et al.* (2) for the two-factor case, and Uzawa (37) and McFadden (27) for the n-factor case.

All the problems show up in the three-factor case, so assume constant returns to scale and $Q = F(X, Y, Z)$. I suppose the most straightforward definition of the elasticity of substitution between X and Y is what McFadden calls the direct (partial) elasticity of substitution: Apply the two-factor definition to X and Y, holding fixed the other factor(s) Z. That is to say, fix Q and Z and thus define a curve in the XY plane which can play the role of a two-dimensional isoquant; along that curve calculate the elasticity of X/Y with respect to $P_X/P_Y = dx/dy$. The formula (4) cannot be applied directly because it involves a use of Euler's theorem, which is of course improper when there are other factors. One does find, however,

$$\sigma_{XY}{}^{1} = \frac{(XF_X + YF_Y)F_XF_Y}{-F_{XX}F_Y{}^2 + 2F_{XY}F_XF_Y - F_{YY}F_X{}^2} = \frac{(XF_X + YF_Y)F_XF_Y}{\begin{vmatrix} 0 & F_X & F_Y \\ F_X & F_{XX} & F_{XY} \\ F_Y & F_{YX} & F_{YY} \end{vmatrix}}$$

This is true without any assumption about returns to scale; if F is homogeneous of any degree, however, $\sigma_{XY}{}^{1}$ depends only on factor proportions. It is clear from the definition, though, that for the competitive constant-returns situation $\sigma_{XY}{}^1$ is the appropriate concept for answering this question: If X and Y are available at given prices while the other factors are fixed in amount, will a rise in p_X relative to p_Y be associated with an increase or a decrease in the ratio of outlays on X to outlays on Y? Or, as the question more often arises in a macroeconomic vein: If the input of X rises relative to the input of Y, all other inputs constant, will outlays on X rise or fall relative to outlays on Y? Note that this is not the same thing as asking about distributive shares, since X and Y are not the only factors and all factors' shares may change; thus the share of X may rise relative to Y but fall relative to the total.

An alternative definition, reducing to the first (and indeed to the standard two-factor σ) when $n = 2$, is Allen's partial elasticity of substitution between X and Y

$$\sigma_{xy}{}^2 = \frac{XF_x + YF_y + ZF_z}{XY}\frac{D_{xy}}{D}$$

where D is the determinant

$$\begin{vmatrix} 0 & F_x & F_y & F_z \\ F_x & F_{xx} & F_{xy} & F_{xz} \\ F_y & F_{yx} & F_{yy} & F_{yz} \\ F_z & F_{zx} & F_{zy} & F_{zz} \end{vmatrix}$$

and D_{xy} is the cofactor of F_{XY}. Under constant returns to scale the first numerator factor is simply $F(X, Y, Z)$. There are several ways of describing the economic meaning of $\sigma_{XY}{}^2$. Standard transformations in the theory of production show that

$$\sigma_{XY}{}^2 = \frac{C}{XY}\frac{\partial X}{\partial p_Y} = \frac{E(p_X X)/Ep_Y}{p_Y Y/C}$$

where the notation Ex/Ey stands for the elasticity of x with respect to y and C is the minimum cost of producing a unit of output. C, X, and Y are to be treated as functions of the factor prices. Suppose p_Y goes up by 1 per cent, output and all other factor prices constant. Then, to terms of first order, unit cost will rise by $p_Y Y/C$ per cent and $p_X X$ will rise by a larger or smaller percentage according as $\sigma_{XY}{}^2$ is larger or smaller than one. Thus if $\sigma_{XY}{}^2$ is larger (smaller) than one, an increase in the price of Y, other prices constant, will increase (decrease) the share of X in total costs (or proceeds). But notice that this is not the same thing as holding all other factor inputs constant and increasing the input of Y; it is the same thing, under homogeneity, when there are only two factors, but not when there are three or more. Which is the right assumption depends on the frame of reference; to a firm or small industry, it is more likely to be factor prices that are constant, though output may change, but to the economy as a whole, under maintained full employment, it is more likely to be input totals.

McFadden has introduced a third definition, which he calls the "shadow elasticity of substitution." In concept it is a sort of hybrid of the two mentioned so far; he applies the two-factor definition—the elasticity of input ratio with respect to marginal rate of substitution along an isoquant—holding fixed the prices of the other factors and the unit cost.

Finally, one can seek a definition which does answer the typical macroeconomic question: What happens to the competitively imputed relative share of X if the input of X rises, other factor inputs held constant and all prices permitted to float. (I have seen the concept used this way in an unpublished paper by R. Sato, and in the MIT lectures of Paul Samuelson.) Note that this is a "one-subscript" elasticity of substitution, which we can call σ_X. A natural formula can be found in the following way. Define the elasticity of derived demand for X, θ_X say, as $-EX/Ep_X$, where Y, not p_Y, is constant. Then, in the two-factor case, it is easily verified that $\theta_X = -F_X/XF_{XX}$ and $\sigma_X = \sigma_Y = \sigma = \theta_X(1 - \eta_X) = \theta_Y(1 - \eta_Y)$. In the three- or n-factor situation, it remains true that $\theta_X = -F_X/XF_{XX}$ and that η_X, the competitive share of X increases or decreases with X according as $\theta_X(1 - \eta_X)$ is larger or smaller than one. One can then define

$$\sigma_X = \theta_X(1 - \eta_X)$$

and know that it gives the right answer to the question being asked.

I have mentioned that the immediate objective of most recent work on the elasticity of substitution has been the search for production functions whose elasticities of substitution are constant for all input bundles. In turn, the objective of this search is added flexibility in empirical work; there can be no other objective, except perhaps curiosity, because there is no reason in principle why elasticities of substitution by any definition should be constant. The results have been discouraging. In the two-factor case, the situation is now well known. The various definitions of σ coincide (no subscripts, because there is only one pair of factors) and the production functions with constant σ are of the form

$$F(X, Y) = (aX^{\rho} + bY^{\rho})^{1/\rho}$$

where $\rho = 1 - \dfrac{1}{\sigma}$.

The work of Uzawa and McFadden has shown that the three- or n-factor case is unrewarding. The natural extension of the two-factor function

$$F(X, Y, Z) = (aX^{\rho} + bY^{\rho} + cZ^{\rho})^{1/\rho}$$

still has constant elasticity of substitution, and the various definitions coincide. The trouble is that all factors (or all pairs of factors) have the *same* elasticity of substitution. That is unsatisfactory, but it seems that very little additional flexibility is possible. For the "direct" and "shadow" definitions the situation is as follows: The factors must be divided into classes; between any pair of factors in the same class, the elasticity of substitution must be unity; and indeed their competitive shares must be identical; for *any* pair of factors in different classes, the elasticity of substitution has the *same* common value. For the Allen partial elasticity of substitution the situation is a little more flexible: Again the factors are partitioned into classes; between any pair from different classes the elasticity of substitution is unity; each pair from the same class has a common elasticity of substitution, the same for each pair from a given class, but possibly different for each class. For the fourth definition given above, the Arrow-Chenery-Minhas-Solow function has constant σ, but necessarily the same for each factor. It is not known if any other function admits constant σ by that definition.

Since the search for three-factor production functions with constant elasticities of substitution has yielded so little, one may seek a wider class of production functions. Mrs. V. Mukerji (28) has observed that the natural generalization of (5)

$$F(X, Y, Z) = (aX^{\rho_1} + bY^{\rho_2} + cZ^{\rho_3})^{1/\rho}$$

has the property

$$\frac{\sigma_{xy}{}^{1}}{\sigma_{xz}{}^{1}} = \frac{1 - \rho_3}{1 - \rho_2}, \text{ etc.}$$

In general, for these production functions, the Allen elasticities of substitution may vary, but their *ratios* are constant. [Note that the only homogeneous functions in this class are the powers of (5).] Something a bit stronger is actually true, namely that $\sigma_{ij}^1/\sigma_{mn}^1 = c_i c_j / c_m c_n$ with the c's constant. But Gorman (16) has proved that, apart from some generalizations to allow for limitational factors and to exploit a little more fully the departure from homogeneity, the proportionality of Allen elasticities of substitution implies that the technology is either of the Uzawa or the Mukerji type.

So if anyone wants to estimate more-than-three-factor production functions to study substitution possibilities, his choice is still pretty limited.

A Theorem of Houthakker's on Aggregation

I have not reviewed any of the recent work on formal aggregation, but I cannot resist mentioning a result due to Houthakker (19), in the hope that someone will take it up and push it further. The assumption is that production within some aggregate, like an industry or even an economy, is carried on in cells, which may be firms or establishments or even places. Within each cell there are fixed factor proportions. Factors are divided into fixed and variable inputs; the variable ones are available to each cell at common fixed prices, while the fixed factors are peculiar to each cell. The fixed factors may represent capital equipment or entrepreneurial ability or locational advantage or anything; they need never by aggregated. Although there are fixed proportions within each cell, different cells may have different requirements for the variable factors. Under competitive assumptions, at any constellation of prices for output and the variable inputs, those cells will produce which can earn nonnegative quasirents after paying the variable inputs, and those cells will produce at capacity. (If there are a lot of cells just breaking even, there will be some indeterminacy. Competitive assumptions can be replaced by any definite alternative assumptions.)

Houthakker's analysis is good for any number of variable inputs; he calculates explicitly for two; I will take the case of one, just to give the idea. Let Y be the variable factor and p_Y its price in terms of product; let y be the requirement of Y per unit of output in an arbitrary cell. That cell will produce to capacity if $p_y Y \leqq 1$, else it will not produce. Imagine there are so many cells, with such fine gradations, that it is reasonable to think of them as a continuum. (This is not essential; indeed for practical work one would want a computer and discrete cells anyway, but integrating is neater than summing.) Let $g(y)$ be the density function giving the capacity of cells with labor requirement y, so that $g(y)dy$ is the capacity of cells whose labor requirements per unit of output lie between y and $y + dy$. At any price p_Y, the aggregate output produced will be

$$Q = \int_0^{1/p_Y} g(y)dy$$

and the total input of the variable factor will be

$$Y = \int_0^{1/p_y} yg(y)dy.$$

Elimination of p_y between these equations yields an aggregate production function giving Q as a function of Y or, in the more general case, of all the variable inputs. In this one-variable-input case, the relation between Q and Y is formally identical to an ordinary Lorenz curve.

For instance, as Houthakker shows, if $g(y)$ is the Pareto-like density Ay^{h-1} (one would want to have h greater than one if one would like to have the density tend to zero with y, but not too much greater or else the integrals would not converge), then the result is Q = constant times $Y^{(h/h+1)}$, which is, of course, a one-factor Cobb-Douglas. The remaining fraction $(1/h + 1)$ of output is imputed as quasirents to the fixed factors.

A slightly different interpretation can be imposed on this structure. Suppose that y has been changing monotonically in time under the influence of technical progress and changing economic conditions; the normal presumption is that $y(t)$ has been decreasing. Suppose that investment, measured in net additions to capacity has been $I(t)$. Then $g(y)$ is found by inverting $y(t)$ to give t as a function of y, and substituting in $I(t)$. If both $y(t)$ and $I(t)$ are exponentials, one comes back to the Houthakker case. It is easy to introduce sudden-death depreciation, or any other simple mechanism for physical mortality. When there is more than one variable factor, this interpretation of the Houthakker procedure allows the desirable property that more than one set of factor proportions be embodied in each year's capacity; different cells may face different conditions or have different expectations.

Houthakker himself treats only the Pareto distribution, which gives rise to the Cobb-Douglas. The calculations can also be carried out with exponential or gamma-type distributions; they lead to a legitimate but not especially convenient aggregate production function. Can anyone think of other interesting cases? Even some numerical calculations would be worth having.

"Putty-Clay" Models

Finally, I want to mention, but not really to discuss, the so-called putty-clay model of production. The name was coined by Phelps to describe technologies in which factor proportions are variable *ex ante,* before capital has been committed to concrete form, but fixed *ex post.* A new nickname is needed; and so is an extension to the (presumably more realistic?) case where some *ex post* variability remains, but with an elasticity of substitution smaller than the *ex ante* one. The idea was pioneered by Johansen (20) in a paper that left aside the value-theory implications entirely. Subsequent work by Phelps (29), Pyatt (30), and Solow (35) has filled in some of the gaps, but there is a lot more to be done. Almost no empirical work has been based on the putty-clay idea: Pyatt's attempt is not very successful, perhaps because it relies on a convenient but unsatisfactory assumption about the choice of technique: that first-year quasirents per unit of investment be maximized.

The importance of the putty-clay model is that it gives prominence to obsolesence—the erosion of quasirents through the competition of newer and more efficient plant. It also poses very sharply some important questions about behavior in the short run: What is maximized, the value of the competitive approximation, the degree of monopoly when aggregate effective demand is deficient. One of the useful functions of theory is the suggestion of new kinds of data it would be interesting to collect. In this model, as I have said, the key concept is the stream of quasirents yielded by a capital investment from the time it is made until it expires either from physical wear and tear or because it can no longer cover prime costs (or some noncompetitive alternative). Can we get such "life cycle" data on revenues and costs?

References

1. Arrow, K. J., "The Economic Implications of Learning by Doing," *Review of Economic Studies,* 29 (1962), 155–73.
2. Arrow, K. J., H. Chenery, B. Minhas, and R. Solow, "Capital-Labor Substitution and Economic Efficiency," *Review of Economics and Statistics,* 63 (1961), 225–50.

3. Black, J., "The Technical Progress Function and the Production Function," *Economica*, 29 (1962), 166–70.
4. Boulding, K., and W. Spivey, *Linear Programming and the Theory of the Firm*, New York, 1960.
5. Chenery, H. B., "Engineering Production Function," *Quarterly Journal of Economics*, 63 (1949), 507–31.
6. Davidson, R. K., V. Smith, and J. Wiley, *Economics, an Analytical Approach*, Homewood, Ill., 1958.
7. Diamond, P. A., "Disembodied Technical Change in a Two-Sector Model," *Review of Economic Studies*, 32 (1965), 161–68.
8. Dorfman, R., *The Price System*, Englewood Cliffs, N.J., 1964.
9. Dorfman, R., P. A. Samuelson, and R. M. Solow, *Linear Programming and Economic Analysis*, New York, 1958.
10. Drandakis, E. M., and E. S. Phelps, "A Model of Induced Invention, Growth, and Distribution," unpublished, July 1965.
11. Fei, J. C. H., and G. Ranis, "Innovational Intensity and Factor Bias in the Theory of Growth," *International Economic Review*, 6 (1965), 182–98.
12. Fellner, W., "Two Propositions in the Theory of Induced Innovations," *Economic Journal*, 71 (1961), 305–08.
13. Fisher, F. M., "Embodied Technical Change and the Existence of an Aggregate Capital Stock," *Review of Economic Studies*, forthcoming.
14. Furubotn, E. G., "Engineering Data and the Production Function," *American Economic Review*, 55 (1965), 512–16.
15. Gorman, W. M., "Capital Aggregation in Vintage Models," *Review of Economic Studies*, forthcoming.
16. ———, "Production Functions in which the Elasticities of Substitution Stand in Fixed Proportion to Each Other," *Review of Economic Studies*, 32 (1965), 217–24.
17. Hahn, F. H., and R. C. O. Matthews, "The Theory of Economic Growth: A Survey," *Economic Journal*, 74 (1964), 779–902.
18. Heady, E. O., and J. L. Dillon, *Agricultural Production Functions*, Ames, Iowa, 1962.
19. Houthakker, H. S., "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis," *Review of Economic Studies*, 23 (1955), 27–31.
20. Johansen, L., "Substitution versus Fixed Production Coefficients in the Theory of Economic Growth: A Synthesis," *Econometrica*, 27 (1959), 157–76.
21. Kaldor, N., "A Model of Economic Growth," *Economic Journal*, 67 (1957).
22. Kaldor, N., and J. Mirrlees, "A New Model of Economic Growth," *Review of Economic Studies*, 29 (1962), 174–92.
23. Kennedy, C., "Induced Bias in Innovation and the Theory of Distribution," *Economic Journal*, 74 (1964), 541–47.

24. Kurz, M., and A. S. Manne, "Engineering Estimates of Capital-Labor Substitution in Metal Machinery," *American Economic Review*, 53 (1963), 662–81.
25. Levhari, D., "Extensions of Arrow's 'Learning by Doing,'" *Review of Economic Studies*, forthcoming.
26. Manne, A. S., and H. M. Markowitz (ed.), *Studies in Process Analysis*, New York, 1963.
27. McFadden, D., "Further Results on CES Production Functions," *Review of Economic Studies*, 30 (1963), 73–83.
28. Mukerji, V., "A Generalized SMAC Function with Constant Ratios of Elasticities of Substitution," *Review of Economic Studies*, 30 (1963).
29. Phelps, E. S., "Substitution, Fixed Proportions, Growth and Distribution," *International Economic Review*, 4 (1963), 265–88.
30. Pyatt, G., "A Production Functional Model of United Kingdom Manufacturing Industry," in *Econometric Analysis for Economic Planning*, Colston Research Society Symposium, Bristol, 1965.
31. Salter, W. E. G., *Productivity and Technical Change*, Cambridge, Mass., 1960.
32. Samuelson, P. A., "Notes on Weizsäcker-Kennedy Theories of Induced Invention," unpublished, 1965.
33. ———, "Simple Intuitive Proofs of Neoclassical Functions Admitting of Capital Aggregation," *Review of Economic Studies*, forthcoming.
34. Sheshinski, E., "Factor-Augmenting, Disembodied, Technical Change," unpublished, 1965.
35. Solow, R. M., "Substitution and Fixed Proportions in the Theory of Capital," *Review of Economic Studies*, 29 (1962), 207–18.
36. Solow, R. M., J. Tobin, C. C. von Weizsäcker, and M. E. Yaari, "A Model of Fixed Capital without Substitution," *Review of Economic Studies*, forthcoming.
37. Uzawa, H., "Production Functions with Constant Elasticities of Substitution," *Review of Economic Studies*, 30 (1962), 291–99.
38. Walters, A. A., "Production and Cost Functions: An Econometric Survey," *Econometrica*, 31 (1963), 1–66.
39. Weizsäcker, C. C. von, "Zur ökonomischen Theorie des technischen Fortschritts," unpublished, 1964.

COMMENT

JAMES TOBIN, Yale University

Professor Solow's survey is a valuable and lucid short course in recent production theory. He claims neither completeness nor representativeness in his coverage. I will not try to judge this claim. But I do think that the topics and contributions that have excited his interest can be taken as indicative of the state of the subject. The generalization which his selective survey provokes is that analytical convenience in related branches of theory has been the main influence shaping the theory of production in recent years. The concepts, assumptions, and formulations which sustain theoretical interest are those which lend themselves easily to getting results in the theory of capital accumulation and growth. The shape of production theory would be different had it been aimed instead at providing a convenient theoretical framework for estimating technological relationships.

The principal exception in Solow's catalogue is activity analysis, a development motivated principally to approximate the complexity of production processes. But its very flexibility and generality seem to confine it to the role of a specific problem-solving technique or to use in abstract models of competitive equilibrium. Anyway it plays little part in the ferment about capital and growth to which so much current work on production is oriented.

Two-factor aggregation. Of the many aspects of theorizing geared directly to this interest, the most prominent is the aggregation of inputs into two factors, capital and labor. (Ghosts from the nineteenth century as, "What happened to land?") One reason for two-factor aggregation is that blackboards are two-dimensional. But the main reason certainly is that growth models focus attention on how these two aggregates differ in the mechanisms determining their supply. Capital is generated by saving from current production, labor by demographic factors usually assumed to proceed exogenously at a natural rate.

This motivation, however, has lost some of its force by recognition that saving can be embodied in human beings, through expenditures on education and health, as well as in physical goods. In any case it is not obvious that classifying inputs by origin should also be the appropriate way to aggregate them in describing the technology of production. That requires various capital goods to be better substitutes for each other, and various kinds of labor better substitutes for each other, than capital goods are for kinds of labor. In a two-factor production function robots would be better thrown with human labor than with floor space, even though they are "produced means of production" resulting from saving.

The embodiment of technical progress in successive vintages of gross investment is certainly a brilliant and seminal idea, for which the profession is greatly indebted to Professor Solow himself. But its appealing simplicity does depend on the aggregation of all the investment of one vintage into one homogeneous productive factor. The model becomes very complicated if a variety of capital goods, obsolescing at different rates, is allowed—plant, equipment, inventories, houses, consumer durables.

Factor-augmenting technical change. The assumption that technical progress augments one or the other input within a stable production function is a powerful simplification, no doubt useful in empirical work as well as in growth theory. But as Solow points out, the approach can be misleading if improvements apparently embodied in a factor are identified as augmenting it. Innovations embodied in new machinery may be labor-augmenting, of course, and the education of farmers may be land-augmenting. The spirit of the Phelps-Nelson approach to the productivity of education—that education enlarges choice of technology—seems more promising than the assumption, made by Denison among others, that it stretches man-hours.

Induced innovation. The notion of factor-augmenting progress is the basis for the new theories of induced innovation reviewed by Solow. The principal interest of these theories, perhaps their principal motivation, is the explanation they give for the stability of distributive shares over time. Unfortunately, although there are plausible versions of these theories which imply such stability, there are equally plausible assumptions which do not. The trouble is that the opportunity locus describing the terms of trade-off for the economy between labor augmentation and capital augmentation is a *deus ex machina.* How does the process work for the individual firm? What explains the concavity of the locus? What scarce resources determine its position? Why cannot it be moved by increasing these resources? Do they get paid, and if so how does their payment affect the theory of distribution?

Learning by doing. As Solow remarks, the Arrow model is most interesting and merits further work. Earlier contributions on the same subject, notably by Hirsch, have related learning to cumulative production. Arrow assumes that we learn only from investment; this indeed is the reason that investment has a higher social than private return. Saving and investment would not carry this extra benefit if production of consumer goods were equally instructive. Here again a priori reasoning cannot choose between plausible assumptions which have quite diverse implications.

Omitted problems. Solow's survey does not report any recent theoretical work in some areas where it is badly needed. I think, for one example, of the theory of depreciation. The assumption of exponential decay contradicts common sense and casual observation. It owes its popularity not to any evidence that this is the way capital goods wear out but to the abundant evidence that it fits smoothly into growth models. The idea that depreciation depends on intensity of use as well as on passage of time is so outmoded that the term "user cost" has been appropriated to mean a cost that does not depend on use at all but only on time.

I find more surprising the omission from Solow's survey of an important challenge which he himself has taken up. This is to provide a theory which will explain input-output relations observed in short-run fluctuations as well as in long-run growth. Perhaps our failure to reconcile these two kinds of observations should lead us, among other things, to question the complacency Solow expresses regarding the basic general neoclassical production function. It is a static function, a relationship among simultaneous steady flows of outputs and inputs. We dodge the difficult problem of specifying the timing of inputs and related outputs by assuming stationary conditions. But we have no right to assume that the relations of outputs to employment and other inputs which would hold when outputs are stationary will also hold when outputs are changing. Nor should we assume, as current growth models generally do, that there is no lag between investment expenditures and the availability of the resulting capital formation as productive input.

I too am greatly interested in growth theory. But I do think that the theory of production deserves a life of its own, with the purpose of providing models which better represent and simplify the facts of technology. This is a worthy purpose in itself, and it may be that pursuing it will also advance in the long run the related theories of capital accumulation and growth.