THE TECHNOLOGICAL PROGRESS FUNCTION AND LABOR FORCE SIZE: A MODEL OF INVENTION FIT TO FACTS

Julian L. Simon, Professor, Departments of Economics and Business Administration

#702
THE TECHNOLOGICAL PROGRESS FUNCTION AND LABOR FORCE SIZE: A MODEL OF INVENTION FIT TO FACTS

Julian L. Simon, Professor, Departments of Economics and Business Administration

Summary

This paper offers a micro model of the relationship of the size of the population and the economy to productivity change. The constituents of the model are the number of potential technology producers, the number of elements available in the environment to stimulate invention, and the probabilistic relationships between inventors and stimuli.

The model works out the tradeoff between the idea-stock-reducing and idea-stock-increasing forces, under different conditions of population growth. The model shows that under assumptions that seem economically and psychologically reasonable, a larger labor force has increasing returns in technology production. The number of possible new combinations that result from the addition of a new idea element to the pool of technical knowledge is very large relative to the depletion of the preexisting pool of technical knowledge by the discovery of that one idea. And the number of such possible new combinations that a representative additional person creates is very large relative to the additional possibilities of duplicated effort that result from the additional person. Hence the net result of an additional person is an increase in the total number of new ideas. And this effect will continue with increases in total persons and total technology until duplication approaches zero, and unduplicated idea production for the representative person approaches the person's total capacity for idea production. Hence there are increasing returns in idea-production to additional persons until convergence to constant returns.
THE TECHNOLOGICAL PROGRESS FUNCTION AND LABOR FORCE SIZE:
A MODEL OF INVENTION FIT TO FACTS

Julian L. Simon*

I. INTRODUCTION

Understanding the mechanism that produces technical progress is a vital task for economic science for several reasons: (1) Our view of what speeds or hinders technical progress touches important policy questions, as for example: Should a country increase immigration? And should it subsidize capital investment? (2) Technical progress is the key issue in growth theory. As Eltis put it in a discussion of equilibrium growth, "How the rate of technical progress is determined will therefore be a matter of the utmost importance" (1973, p. 129). (3) It is also the most important question in the economics of population—the underlying interest of this paper—because it determines the effect of additional people on the standard of living of the community.

Given the importance of the topic, the paucity of attention to it is surprising. Furthermore, the bulk of the modern attention has gone into the question of the direction of advances in technique, whether

*Mark Browning, Jim Smith, and Gunter Steinmann gave me enjoyable conversation and useful ideas on the subject of this paper. Fritz Machlup read the paper and provided valuable criticism. At his behest, I shall strive for precision in the use of the following terms: "technology" or the redundant "technological knowledge" shall mean the body of existing knowledge about techniques that are useful for economic purposes, and "technological progress" shall mean the increase in technology. "Technical progress" shall mean improvement in techniques used, and the increase in the level of technology actually at work in the economy. I shall avoid the use of the general word "knowledge" which, in Machlup's classification, includes both technology and all kinds of other knowledge including "spiritual, intellectual (by which [he means] useless), practical and positive entertainment knowledge" (Correspondence, December 17, 1979).
affected by relative intensities of capital and labor, and whether influenced by demand in particular industries or autonomous (see, for example, the 1962 NBER volume edited by Nelson). But the more important question from the point of view of both policy and growth theory is the determinants of the total quantity of technical progress and the rate of change in the level of technique.

Discussion of induced technical progress has been almost entirely at the level of the economy, the industry, or the firm. And that makes considerable sense because the available data are at those various levels of aggregation, ranging from the Abramowitz-Solow-Denison tradition of work on the components of growth, to the Rostas-Verdoorn tradition of work on industries, to the Alchian-Arrow learning-by-doing tradition, and the industrial-organization and organizational-behavior literatures on firms. But there has been little discussion and less formal analysis of induced technical progress at the level of the individual inventor and adapter. (I shall refer to an analysis of invention at this level as a micro-model, or a model of invention, to be distinguished from both a more aggregated analysis of technical progress, and from a macro-model of the economy as a whole that embodies a technical progress function.) Without a model of the actual actors, and their interactions with each other and with the body of existing knowledge, we are talking about a mechanism taking place in a black box, proceeding without any understanding of its mode of operation.

Furthermore, the technical progress functions that have been proposed in the literature—Kaldor's, Arrow's, Phelps', and others—vary greatly in their implications, and they would not be consistent with the
same micro-models. Therefore, a micro-analysis of invention can throw some light upon which of the contending technical progress functions and accompanying macro-models of the economy is most plausible.

An additional complication in the chain from invention to its effect on productivity and the standard of living is the relationship between invention and adoption of new techniques. At any one moment there is a large stock of useful techniques that have not been applied, and relatively few inventions ever get used—1 in 1000, in Machlup's guesstimate. And the relationship can be variable, with diminishing returns to additional inventions in the short-run. But in a long-run context, the number of potential appliers of knowledge should be proportional to the number of inventors, and therefore, I shall assume that technical progress is a proportional function of advance in technology, without further discussion of the relationship between them.

The aim of this paper, then, is to offer a model of the process of technological knowledge creation at the individual level, to see what it implies and how it fits with the aggregate evidence and the macro-models. Of course such a model is not likely to persuade everyone, or even most people (though hopefully it will persuade some) that it is reasonable. And no one at all will think that by itself this model is adequate, or even more than a rough beginning. But it is hoped that this model will serve to initiate systematic discussion that will eventuate in a satisfactory micro-model of the process of technical progress, especially in relationship to population growth.

Section II discusses the main theoretical consideration that may influence the relationship between technical progress and labor-force
size and growth. Section III discusses how various macro-models in
growth theory have dealt with technical progress. Section IV constructs
a micro-model of invention relating population to advance in technology,
and considers how it fits the various macro-models of economic growth.
Section V discusses the results and qualifications of them. Section VI
concludes and summarizes.

II. ECONOMIC SPECULATION ON THE DETERMINANTS
OF THE TECHNOLOGICAL PROGRESS FUNCTION

The earliest economic speculation on the source of technological
progress known to me is that of Petty, who emphasized the number of po-
tential inventors:

"As for the Arts of Delight and Ornament, they are
best promoted by the greatest number of emulators.
And it is more likely that one ingenious curious man
may rather be found among 4 million than 400 persons
. . . And for the propagation and improvement of use-
ful learning, the same may be said concerning it as
above-said concerning . . . the Arts of Delight and
Ornaments . . ."  (1632)

That is, if there are more potential knowledge creators and more minds
at work, there will be more practical ideas created and adopted. In
1960 Kuznets added important substance to this brief remark of Petty's.

Though many of the technological progress functions that we will
discuss were originally written in terms of the rate of change of tech-
ology, it seems more concrete and therefore clearer and easier to think
about a function that has as a dependent variable the absolute numbers of
inventions and discoveries, rather than a more abstract rate of change
which is itself a comparison of two magnitudes.* We therefore express the Petty-Kuznets idea as

\( A_t - A_{t-1} = f(L_t) \)

where \( A_t = \) the level of technique in use at time \( t \)

\( L_t = \) labor force, considered here as proportion to total population.

There is broad support for this formulation in the fact that there has been more progress in those centuries and millenia when population was greater—say the last millennium or century, compared to periods two thousand years ago; the facts for the U.S. also show an increase in the rate of change of productivity for this century, during which data are available (Solow, 1957; Fellner, 1970, pp. 1112; for a summary see Simon, 1977, Chapter 4).

A more general view is that not only people but also income and a stock of technology are important factors of technological knowledge production. The latter matters because ideas build upon ideas. The former matters because technology production clearly is a function of the richness of the economy, either because of the expenditures on education or because of the variety of objects and situations people meet up with, or—most plausibly—both. So we write

---

*The rate-of-change ratio involves two quantities and a comparison of them, rather than just one quantity, which is why I believe that we would make faster progress if the discussion focuses on the absolute quantities. As Leontief put it: "[I]n the actual process of scientific investigation, which consists in its larger part of more or less successful attempts to overcome our own intellectual inertia, the problem of proper arrangement of formal analytical tools acquires fundamental importance." (1966, p. 59)
\begin{equation}
A_t - A_{t-1} = f[L_t, A_{t-1}, \left(\frac{Y}{L}\right)_{t-1}]
\end{equation}

where \(Y\) = national income.

From the point of view of general policy questions, as well as social decisions about population growth, the key issue is not that of deciding which sort of input factor—investment, output, or labor force, is the most important. Rather, the key question is whether any measure related to the size of the economy or the population has a strong effect on productivity. This behooves us to ask: Under which conditions might additional people not lead to increased productivity?

1. Additional people dilute the stock of capital, even if they increase the stock of technology. Therefore, the trade-off between these two forces must be studied in the context of a full economic model.

2. Some forces involved in the creation and adoption of technology are said to be "inversely related to the rate of population growth (Spengler, 1968, p. 115)." A prominent example: Faster growth and a larger population might lead to less education per person. But the data (Simon and Pilarski, 1979) suggest that this force does not operate strongly in the short run. And in the long course of history so far, this surely has not been much of a factor. Another possibility is that faster population growth reduces the rate of saving, and hence reduces the rate at which technology, embodied in capital, is brought into use. But the evidence on the population growth-saving relationship is very thin.

3. Another possibility is that additional people are so much like existing people that they are not likely to come up with any improvements
that the existing population would not come up with. That is, people may be sufficiently homogeneous so that additional people do not produce additional variety in ideas. Evidence for similarity in people's thinking is found in the existence of independent inventions of the same idea, e.g., stochastic dominance is a recent example in finance, modeling of the consumer is an example in marketing, and DNA an example in biology; patent priority fights are additional evidence. This possibility implicitly assumes a limited base of existing technology which people can develop, and a limited range of other stimuli to people's imaginations.

4. If there is a very obvious order in the value of potential projects, if people are clear-minded enough to perceive this order, and if there is at least a fair amount of similarity in people's talents and interests, then there would be high duplication in work on the most important inventions.

Surely there is some added variety introduced by additional people. Therefore, even if there are diminishing returns in technology production, in the very longest run an increment to technology now will have a more positive effect on income than any decrement to capital per worker now due to additional people, on almost any assumptions. The question, then, must be: How much additional productive knowledge can we expect from additional people, and how long will it take to overcome the negative effects of capital dilution?

Furthermore, there are reasons to suppose that in the long run there are increasing returns to additional persons even in technology production itself. Kuznets makes an argument for increasing returns on two grounds: (1) the stimulating effect of a dense environment; and
(2) "interdependence of knowledge of the various parts of the world in which we human beings operate," (p. 328) e.g., discoveries in physics stimulate discoveries in biology, and vice versa. And he discounts the possibility of diminishing returns because "the universe is far too vast relative to the size of our planet and what we know about it" (p. 329).

Machlup suggests that "every new invention furnishes a new idea for potential combination with vast numbers of existing ideas . . . [and] the number of possible combinations increases geometrically with the number of elements at hand" (1962, p. 156). It is this latter idea of an increasing number of permutations of elements of knowledge as the stock increases that seems most compelling to me, when put together with the idea of a reduced possibility of duplicate discoveries as the number of possibilities increases faster than the number of potential technology producers. And we shall see later that—contrary to intuition—duplication of intellectual effort is likely to be less and less of a constraint with the passage of time and with larger populations. Eventually we can expect the main constraint to be an individual's idea-production capacity alone, while the duplication factor will be negligible.

Compelling evidence on the contribution of additional people would seem to be found in the aggregate evidence on productivity—inter-country and intra-industry comparisons, as well as learning-by-doing. It seems inescapable to me that this implies that more persons lead to more technical progress through more individuals producing technical knowledge. But many others are quite unwilling to accept this as proof that more workers raise productivity. Therefore we must investigate the process at the individual level, to build a model and to speculate about reasonable parameters to see how they jibe with the aggregate evidence.
III. TECHNICAL PROGRESS FUNCTIONS IN GROWTH THEORY

Most economists have assumed—explicitly or implicitly—that technical progress is in fact independent of the size of the economy and the population.

As Ansley Coale put it:

[T]here is no warrant for the assumption that growth and knowledge is greater with a larger population.

... I think even the most cursory consideration of scientific or more general cultural history would bring to light too many counter examples to make this theory tenable. I have gifted and well-informed friends who seriously think that the intellectual heights achieved in classical Athens have never been equalled, and this was a community of a few thousand educated persons. The population of Florence at the time of the Renaissance was no greater than Trenton, New Jersey, yet Galileo was one of the key figures in the development of modern science; the Medici and their fellow bankers were pioneers in the development of modern banking, including double entry book-keeping; Dante is a figure in world literature rivalled only by Shakespeare and possibly Homer; Machiavelli is considered by some the godfather of political science, and in painting, sculpture, architecture and engineering the Florentines led the world. One could plausibly argue that this community of a hundred thousand persons did more for modern civilization in a few centuries that the U.S. has. Elizabethan London and Budapest between the two world wars (in fact, the Jewish community in Budapest) are other examples.

(Correspondence, December 28, 1971).

For years this assumption was made operational by simply omitting technical progress from mathematical discussions of growth. When technical progress was finally introduced explicitly, it was brought in as a constant function of time. And in the context of the discussion of population growth almost no one says that greater population should mean faster
technical progress* by the action of additional minds—even though it is a well-accepted fact that productivity increases as a function of total output through learning-by-doing, and though productivity obviously is affected by basic science and R&D activity in the past.

In 1949 Verdoorn speculated that "a change in the volume of production . . . tends to be associated with an average increase in labour production" (1949-1979, p. 1), division of labor being the process he had in mind. And he showed a long-run historical relationship in various countries between the rate of increase in total product and the rate of increase in productivity. Verdoorn's technical progress function is

\[
A = ay
\]

or

\[
\frac{A_t - A_{t-1}}{A_{t-1}} = a\left(\frac{Y_t - Y_{t-1}}{Y_{t-1}}\right)^{\frac{1}{2}}
\]

where \(A\) = the level of technique
\(Y\) = gross national product
\(\cdot\) = a dot indicates a rate of change
\(a\) = constant.

But Verdoorn did not deal with the question of the direction of causation, that is, whether faster technical progress causes faster economic growth rather than (or in addition to) the converse. Verdoorn's function also has the unlikely property of implying that if there is zero economic growth

*Even Milton Friedman, whose perception of basic economic forces is usually keen, did not accept Kuznets' proposition that more minds mean faster growth of productivity, when commenting on the 1960 article in which Kuznets advanced this idea (Friedman, 1960, "Reply", pp. 349-350).
growth there would be zero technical progress. In a definitional statisti-
cal sense this may well be true. But behaviorally we can be sure that some invention and innovation occurs even in the absence of measured economic growth, as seen in the famous Horndal effect.

Kaldor (1957) picked up Verdoorn's notion, dubbed it a "law", and introduced it into his growth theorizing. But in his writings he turned from output to capital investment as the determinant of technical progress. His original function was:

\[ \dot{A} = a(K), \]

where \( K = \text{capital} \)

and from it Kaldor concluded (incorrectly) that population growth would have a neutral or negative effect on the growth of income per capita. He also made no note of a possible positive effect of a larger total population through the Petty effect. And as with Verdoorn's function, if investment is zero, technical progress is zero, an unlikely outcome.

Kaldor (with Mirrlees) later worked with the function

\[ \dot{A} = a\left(\frac{K}{L}\right) \]

That is, "the annual rate of growth of productivity per worker...[is] a function of the rate of growth of investment per worker" (Kaldor and Mirrlees, 1962/1970, p. 309). This function shows population growth and size to be even less favorable than (4).

Eltis' technical progress function differs somewhat from Kaldor's, being (1973, p. 151)
As does Kaldor's, Eltis' function has the property that though it varies over the business cycle, it implies a constant long-run rate of change of technical progress,* which may or may not be reasonable, as we shall discuss later. An unrealistic feature of Eltis' model is that it comprehends only embodied technical progress (p. 18); disembodied technical progress must be at least as important.

Alchian took note of the fact that production of such products as airframes improves in the course of production. He then introduced this "learning-by-doing" insight into economic theory, distinguishing among various sorts of economies of scale (1949/1963; 1959). Arrow built an explicit technical-progress function upon this foundation. Like Kaldor he shifted from output to capital as the carrier (or the proxy for the carrier) of embodied technical progress, but unlike Kaldor's, his capital stock variable was intended to be a proxy for cumulative output:

\[ A = aK^b \]

where \( b \) is a constant analogous to the coefficient of serial numbers in learning-by-doing studies, and is of the order of .2. Arrow's technical-progress function has a variety of defects for the understanding of the effects of different population sizes and growth rates, however, some stemming from the inherent ambiguities in the capital concept, some stemming from the alteration in meaning with the shift in variables from

*If profits decline, however, the rate of progress can increase, in Eltis' model.
output (the empirical basis for the learning-by-doing phenomenon) to capital; these matters are discussed in another paper (Simon, 1979a). Also, as with Kaldor's model (which Arrow calls similar to his own), stable production implies no learning-by-doing and no technical progress, which Arrow himself notes is contrary to the Horndal effect and similar evidence elsewhere.

Shell (1966) wrote a function in which the stock of technical knowledge in use is a proportional function of current output (less the decay in knowledge, which can be disregarded here)

\( A_t = a Y_t \)  

The constant \( a \) here reflects both the R&D level and the industry success rate of inventions. This function suffers from several drawbacks, including the mutual influence of \( A \) and \( Y \), and the implication that \( A \) will decline if \( Y \) declines and will not increase if \( Y \) is constant. Putting absolute \( A \) rather than a change in \( A \) on the left side does not seem a promising way to model the invention process at the micro-level.

An attractive function is that of Phelps (1966)

\[ A_t = \frac{A_t - A_{t-1}}{A_t} = h \left( \frac{L_{t-1}}{A_{t-1}} \right) \text{ where } h \text{ is concave.} \]

This function has an increasing \( A \) even with stationary population, which is realistic, and it has steady-state properties, which makes it esthetically pleasing. It has a constant \( A \) with constant \( L \), or decreasing \( A \) with decreasing \( L \), (the latter running contrary to Western experience in the 20th century). Phelps' model also has the property—unmentioned by Phelps—that the rate of growth is a monotonic
positive function of population growth; this is a startling contradiction of all conventional growth theory.

The models described above are summarized in Table 1, where we see that they differ from each other and from the historical facts in a variety of ways. A sound micro-model of invention should help us decide which technical progress function is most to be preferred.

### Table 1

**IV. A MICRO-MODEL OF INVENTION**

At any moment there are \( L_t \) workers in the economy, all of whom are potential creators of new technical knowledge. These workers differ in their propensities to invent and innovate, and one could differentiate them in the model. But that would introduce complication without changing the conclusions.

Let us first notice that no matter what assumptions one makes about the composition of the "original" population—about the homogeneity, and about the distribution of characteristics—an increment of people similar in composition to the original people will increase the number of ideas by that same proportion if the original group and the incremental group are exposed to different but equally-potent stimuli. That is, if people and stimuli are all that go into idea-making, the function is homogenous of degree one. There may be individual sluggards and individual big producers, but each kind will be found in each group. Critical-minimum-size groups may be necessary to produce some kinds of ideas, but this will be true in the same way for the original as for incremental groups.
Characteristics of Technical Progress Functions

| TABLE 1 |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | Knowledge      | Embodied        | Both            | Technological   | Increasing      | Rate of M      | Perhaps         |
|                | Embodied       | Embedded        | Both            | Technological   | Increasing      | Rate of M      | Perhaps         |
|                | Not            | Not             | No              | Yes             | No              | Yes            | No              |
|                | Yes            | Yes             | Yes             | Yes             | Yes             | Yes            | Yes             |

Consequent parameters: Arrow (Verdoorn Kaldor-Kaldor), Arrow (Verdoorn Kaldor-Kaldor), Arrow (Verdoorn Kaldor-Kaldor), Arrow (Verdoorn Kaldor-Kaldor), Arrow (Verdoorn Kaldor-Kaldor).
In brief, the only constraint or factor that might lead to diminishing returns to additional people in idea production (leaving aside education and capital for now) is the size of the stock of knowledge that is available in common to the potential idea producers. This means—somewhat surprisingly—that we do not need to know anything about the degree of similarity of additional people to existing people in order to know how additional people will increment the flow of ideas.

New technology is built upon the stock of technology and other stimuli, i.e., the number of bits of information \( H_t \), in the environment at moment \( t \). This is the key idea and the fundament of this analysis. And there are two relevant aspects of the stock of technology that may be used for creating new ideas. On the one hand, when an idea is created, that idea is removed from the universe of possible new ideas still to be discovered. And if ten more people come along and re-discover the idea independently, nothing is gained; here we find the operation of a process of diminishing returns to additional ideas.* On the other hand, newly-discovered ideas also add to the stock of elements that may be combined with other elements to create still other new ideas. Surely this describes the history of human economic and intellectual growth. Newton's and Einstein's, Smith's and Malthus' and Keynes' discoveries depleted the stock of those potential ideas, but opened up vast fields

*Even when there is duplication, the result is not pure waste. The double exposure increases the chances of the idea being accepted and put into use. And the two discoveries are likely to reveal somewhat different aspects of the phenomenon.
of possibility for future discoveries.* Discoveries about indexes and national income accounts removed those fish from the pond of potential discoveries in economics, but left their eggs to spawn large numbers of new fish which could not have previously hatched. The creation of new economic data—a result of more people and more income in a country—removes the opportunity to do this for the first time, but greatly enhances the possible number of other studies a contemporary economist can do. The research that required brilliance of Adam Smith or John Graunt or William Petty can now be done routinely by us. And an article by Arrow or Becker or Markowitz or Schultz or Stigler can—fortunately or unfortunately—spawn a career industry.

Let $H_{jj}^{\tilde{j}j}$ be an idea formed of the combination of two different stimuli, $j$ and $\tilde{j}$, coming from the set $j = 1, 2, 3, \ldots, m_t$ of stimulus elements existing as of $t$. To be a new idea, this combination $jj$ must never have occurred before, which also means that it is not itself found as an element in set $j = 1, 2, \ldots, m_t$. (A fuller treatment would also consider the higher order combinations, $H_{jj}^{\tilde{j}j\tilde{j}}$, but here we may neglect them.)

The stock of technology available at any moment consists of the stock of technology at the beginning of the prior period plus the new ideas created in that period, $\sum H_{t+1} = \sum H_{t, jj}^{\tilde{j}j} + \sum H_{jj}^{\tilde{j}j\tilde{j}}$ plus all $H_{jj}^{\tilde{j}j}$ newly appearing in $t$. And each new idea is thereby added to the stock of stimuli

---

*In a recent discussion of a nuclear power plant emergency, Royster commented, "The experience should be humbling. It should give us pause. It should remind us of an eternal truth; the more we know, the more the mystery, the more we have to learn." Wall Street Journal, April 11, 1979, p. 22.
available as building blocks for idea-creation in subsequent periods. Notationally, each \( H_{jj} \) occurring in \( t \) becomes a \( j \) in \( t+1 \).

We shall assume that during each period, each person \( i \) among the persons in the labor force \( L \) produces exactly one combination of \( j \) and \( j \), a random drawing without replacement (for that drawing) from \( L_{jj} H_{t,j} \).

In reality, people differ considerably in the numbers of ideas they create, but (as discussed earlier) this will be seen not to affect the model if we notice intuitively that a partition of the labor force into the same proportional divisions each year of idea producers (one idea) and non-idea producers (zero) would not affect the model (nor would a more continuous distribution).

It would also be realistic to assume that the flow of ideas per representative person increases over the years, both because more education likely leads to more productive ideas, and also because there is, to my knowledge, no evidence that scientific production is declining in output per person.* And we see ever-increasing flows of new products, and of such additional carriers of new information as magazines and journals.

Our ultimate aim is to learn how the number of potential idea-producing persons (the labor force) affects the growth of the stock of technology. For economic purposes we will want to know whether the additional ideas produced by additional workers will be sufficient to

---

*This does not apply to the output of a given person over his or her career, but rather to a comparison of similar persons at the same ages in successive cohorts. For a recent review of relevant data in the physical sciences, see Rescher (1978).
counterbalance the negative effect of additional persons on the supply of capital per person.

The number of ideas that might be discovered at time $t$ is

$$M_t = m(m - 1) - \sum_{t=\infty}^{t-1} \sum_{j=1}^{j=m} H_{t,jj}$$

and we assume that if a person first hits on an idea that has previously (not concurrently) been discovered—the negative part of the left-hand side of the above expression—the person will simply "throw the fish back" and find another hitherto-undiscovered idea, but one which may be concurrently discovered by another person. $M_t$ is a simplified symbol for this concept.

The probability that a combination $jj$ is a new idea is the proportion of all possible $m(m-1)$ combinations that have not previously been discovered

$$P(jj) = \frac{m(m-1) - \sum_{t=\infty}^{t-1} \sum_{j=1}^{j=m} H_{t,jj}}{m(m-1)}.$$  

The probability of duplication or "overlap"—the probability that a given new $jj$ will be discovered by more than one person in the same period—is roughly the ratio of the number of other people in the "discovery" labor force, $L_t$, to the total number of possible undiscovered ideas, $M_t$, or $L_t/M_t$. Therefore the total number of new ideas that will be discovered in a given period, not counting duplicates or higher-order overlaps more than once, is

$$L_t \left[ \frac{m(m-1) - \sum_{t=\infty}^{t-1} H_{t,jj}}{m(m-1)} \right] (1 - \frac{L_t}{M_t}) = \sum_{j=1}^{j=m} H_{t,jj}.$$
The total stock of ideas $M_t$ at the end of period $t$ is

$$M_t = \sum_{t=0}^{t-1} \sum_{jj} H_{t, jj}^t + \sum_{jj} H_{t, jj}^t$$

Intuitively we can see that the proportion of overlap will diminish each year because in successive years each new idea-element can combine with a larger number of existing stimuli-elements, as Machlup (1962) noticed. A bit more precisely, the number of elements that is added to the set $m$ each year is almost $L_t^t$, and therefore the number of new possible combinations is almost $L_t^m$—or better, $[L_{t-1}^t + L_t^t]m_t$, where the dot indicates the rate of growth, whereas the number of possible overlaps rises only from a bit less than $L_{t-1}^t$ to a bit less than $L_t^t$, roughly $L_t^t$. If $L_t^t$ is 2% of $L_t^t$, then $L_t^m$ is 50$m_t$ as large as $L_t^t$, a very large factor of multiplication no matter what number—-a million, a thousand, or a trillion—one assigns to $m_t^t$. Given this very large rise in the ratio of to-be-discovered elements relative to the number of potential overlaps, in a relatively short time the number of ideas per period will approach the number of persons, as $\sum_{jj} H_{t, jj}^t + L_t^t$. We know this is not in fact true because people do not work on ideas across the spectrum of possibilities, but cluster where the "action" is. But if the size of cluster remains the same from year to year, we can replace "idea" with "cluster" in our thinking without changing the results.

This analysis implies that there are increasing returns in technology creation from additional people, until the process tapers off to constant returns. This implies that additional people imply additional technology, without limits. And because an increment of technology has
a cumulative effect in raising income through its effect on output and capital formation, a proportional increase in technology in $t$ comes to have a more-than-proportional positive partial effect on income in $t + x$. *

V. THE MICRO-MODEL OF INVENTION EMBEDDED IN A GROWTH MODEL OF THE ECONOMY

There is nothing in the productivity-increasing micro-model that is a counter-balancing disadvantage to the clear tendency for a larger labor force to produce a faster rate of productivity increase. But if we embed the micro-model into a macro-model of the economy, we find that capital dilution works in the opposite direction, more workers implying less capital per worker, ceteris paribus. Therefore we must construct and manipulate the combined model in order to determine how the tradeoffs operate under various conditions.

We begin with the simplest conventional Cobb-Douglas production function:

$$Y_t = A_t L_t^\alpha K_t^\beta, \quad \alpha + \beta = 1$$

The labor force grows at a rate that will be varied for experimental purposes.

*I have implicitly been assuming that technology is permanent and does not obsolesce. To some extent technical knowledge does obsolesce, as we know it does from decreases in use of journals and books with the passage of time (e.g., Fussler and Simon, 1969; but note that some old technical knowledge remains in use by way of being embodied in newer ideas, an effect which does not show up in readership statistics, and from the diminution of patents kept in force (Nordhaus, 1969; but notice that these patent-maintenance data also simply show patents which were finally proven to have no use and be unprofitable, and hence are dropped). To this extent, the analysis must be modified. Obsolescence would seem to make additional people relatively more valuable because the essence of technology production is its external effect. But this needs further study.*
(14) \[ L_t = L_{t-1} + dL_{t-1}. \]

Net saving may be treated as a fixed proportion of output

(15) \[ K_t = K_{t-1} + kY_{t-1}. \]

The micro-model may be built in by equating the level of technique \( A_t \) with the stock of technology \( M_t \). Simulations with a wide variety of parameters, and a technical progress function with specifications consistent with the model given above—for example, \( M_t - M_{t-1} = f(L^\phi, M^\psi) \) or \( = f[L^\phi M^\psi (\frac{Y}{L})^\Delta] \) or \( = f[L^\phi (\frac{Y}{L})^\Delta] \)—all show that higher population growth yields a higher present value of future consumption at discount rates at least up to 6% (to be compared to the historical 2-3% pure rate of discount); the details are given in another paper (Simon, 1979b).

VI. COUNTER-ARGUMENTS AND QUALIFICATIONS

1. Not all ideas that might be discovered have the same potential quality and value. If people came upon ideas only by chance, this non-equality would not in any way affect the workings of the model given above. But people prefer to produce high-value ideas rather than low-value ideas. If it were also the case that people could also identify the possible high-potential ideas with perfect accuracy, then—if there were no differences among people in interests and skills—every potential technology-producer would be working on the same potential idea at a given moment. If there were also a reasonably high likelihood that a random person would succeed in making the discovery of this highest-value idea, then diminishing returns to additional persons would be sharp.
To obtain a satisfying understanding of this mechanism would require a fuller model. Such a model would not be easy to construct. More important, however, qualitative estimates of the relevant parameters are, to my knowledge, both non-existent at present and extraordinarily difficult to produce in the future. The best that we can do at present is to discuss the matter qualitatively and to try to compare the overall results against the aggregate facts, which I will do in that order.

Diminishing returns due to concentration of efforts on the same research project requires, first of all, that all persons have the ability to work in the same general area, and that intellectual mobility is total. But this is clearly not the case. The mature individual who shifts from one major area to another (e.g., from economics to biology or psychology, or the reverse) is rare, and even shifting between minor areas (e.g, public finance and population economics) is not very common. Furthermore, there is proliferation of areas within which people study and later work, so there may not even be a greater number of effective competitors than in the past; for example, Plato and Aristotle, and Bentham and Hume and Smith, took much of human knowledge as their playground, talked to others interested in all these subjects, and made fundamental contributions in all of them; this sort of behavior is much less likely today. It might even turn out (if one were to examine membership lists of professional organizations over hundreds of years) that one's primary professional organization has grown no bigger, e.g., the AFEA or the PAA now versus the AFA at the turn of the century. If the size of "non-competing groups" has not grown over the years, this would be enough to
dispatch the worry of diminishing returns from increasing concentration on the idea with the highest potential value.

Second, diminishing returns from increasing concentration of workers requires ability among workers in a given field to spot the highest-potential projects. Such perfect prevision is far from the rule; even the most eminent men disagree on which areas are most worthy of attention. Furthermore, there is ample evidence that many important ideas are not come upon even though the intellectual preconditions and the need for the idea are present. One example is dynamic programming, which could have been developed and used centuries before it was invented. Another example, according to Weinberg (1977, p. 6), is the set of "discoveries of the recession of distant galaxies and of a weak radio static filling the universe. This is a rich story for the historian of science, filled with false starts, missed opportunities, theoretical preconceptions, and the play of personalities."

Third, people would have to be willing to practice a strategy of working on the project with highest potential but also greatest competition. If researchers are anything like Hotelling's spatial retail competitors, or Downs' political competitors, or several theorists' television-program competitors, they will recognize that it can make sense to move away from the highest-potential area of the market in order to face less competition. Certainly we see this in some areas of science: India and Brazil are of greater importance on the world scene than are Ceylon and Guyana, but some students (even those from the US or USSR) work on the smaller rather than on the larger of them.
With regard to the first three points above, it should be noted that we don't know much about how people pick topics to work upon. Some are attracted by large concentrations of workers; others are attracted by empty areas. Some are influenced by the scientific literature; others—like Keynes in the 1930's—are influenced by the news and the events of the day. Some are influenced by ideas, others by physical objects. Some people like low-risk projects, others are less risk averse. About all we safely say for sure—but this is important—is that there is great variety in people's topic-choosing behavior, as may be seen from the fact that seldom will two members of an academic department be working on the same problem with the same tools and for the same purpose.

Fourth, concentration of workers on a given problem is only a cause of diminishing returns if the likelihood of any one person being successful is great enough so that the likelihood of more than one being successful is meaningfully large. And if this were really the case, then simultaneous duplicate production of important ideas would be seen frequently; it is not, in my observation.

Fifth, for concentration to be a cause of diminishing returns, communication among workers in the same general areas would have to be sufficiently slow so that a person who might work on a potential discovery would not be warned off by the just-previous success of another. All of us have had the experience of contemplating working on an idea, checking the literature and the field of allied workers, and finding out that in fact it has been done. This is not counter-productive duplication, because the resources devoted to the initial insight are
relatively small, and can often be redeployed on the basis of the knowledge of what the other person has already discovered. Given that the speed of relevant scientific communication is almost surely increasing, safeguards against duplication of work on the highest-potential ideas (as well as lower-potential ideas).

We are only interested in the above set of possibilities if all of them are each present to high enough degree to together cause increasing duplication of effort with growing numbers of people. Therefore, an aggregate test is fair: The numbers of persons eligible to make discoveries have been increasing. But the rate of duplication of discovery of the greatest ideas shows no clear direction toward greater duplication. The Leibniz-Newton conjuncture is the last great incident I know of; and, has anyone argued that the circumstances surrounding the discovery of DNA indicate that too much talent was devoted to the task?*

The key issue here is the concavity of \( f(M) \) or \( f(M,L) \) or \( f(M,L,\frac{Y}{L}) \), upon which the number of duplications throw light, but there seems to be no increase in that concavity.

Another relevant test is the rate of investment in R&D: The very long-run trend is up rather than down, even in percentage-of-GNP terms as well as in absolute terms. Of course this might reflect increasing supplies of potential investors. But it is also consistent with an increase in profitable research opportunities, and hence with decreasing duplication.

*It seems as if one could profitably classify discoveries into those that were foreseen by many (e.g., DNA) and those that were not (e.g., relativity). Whether this is so, and what it might imply, is not obvious.
Still another line of evidence refers to the number of first-rate profitable opportunities for research, rather than to the rate of duplication itself. If the number is very large, then duplication of top possibilities is not a problem. Here are a few quotations that indicate the possibilities lying in wait for the idea-producer:

Cancer researcher Cole: "It's not so much a question of more money to look at the problem, as a lack of people and analytical tools for the job." (Blakeslee, 1979).

Newton: "I do not know what I may appear to the world; but to myself I seem to have been only like a boy, playing on the sea-shore, and diverting myself, in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me." (quoted by Taylor and Wheeler, 1966, p. 187).

Bethe (on nuclear fusion): "Money is not the limiting factor: the annual support in the U.S. is well over $100 million, and it is increasing steadily. Progress is limited rather by the availability of highly trained workers, by the time required to build large machines and then by the time required to do significant experiments." (Bethe, 1976, p. 2)

It may also be useful to examine the lengths of time between discoveries that seem equal-importance landmarks in a given field. For example, one could list and measure the times between Ptolemy, Copernicus, Newton, Lorentz and Einstein. The decreasing intervals suggest that great discoveries in physics were not becoming increasingly hard to make.

When thinking about diminishing returns in technology production, it is important to distinguish whether the subject is technology in the small or in the large—that is, advances in technology with respect to a narrowly-defined topic, or technology pertaining to productivity in a wide field or in the economy as a whole. Consider the interesting data in Figure 1. There we see that the rate of technological progress with
respect to any given type of particle acceleration has been much slower than the rate of technological increase in all types of particle accelerations taken together. Clearly there would be much more sharply diminishing returns to additional researchers working on any one type of acceleration than with respect to particle generators as a set. And there would be even less diminution of returns to additional workers in nuclear physics in all its breadth.

**Figure 1**

2. The subject of the discussion so far has been the quantity of technology production. The reader may wonder whether the quality may decline even if the quantity does not. This is a difficult question. It may even be logically impossible to determine the value of later discoveries relative to earlier ones, because the later discoveries depend upon the existence of the earlier ones. This implies that part of the value of the later discoveries should be attributed to the earlier ones, and there is no meaningful way to make this partition. It is like the steps in a ladder: One can put a value on the benefits rendered by the first rung along, but part of the benefit gained from the use of the second rung must be attributed to the first rung.

One possible avenue of inquiry is judgmental: We may ask questions such as: Was Einstein's contribution less valuable than Newton's? Newton's less than Ptolemy's or Archimedes? Some of this has implicitly been done in the lists that have been made of the "great" discoveries by century (see the summary in Sorokin, 1937). But of course such lists
Figure I

ENERGIES ACHIEVED BY PARTICLE ACCELERATORS FROM 1932 TO 1968

Note The exponential growth of accelerators was first noted by John P. Blewett in an Internal Report of the Cosmotron Department of Brookhaven National Laboratory, dated June 1, 1950. The first public presentation of this material was made by Fermi in his address as retiring President, at the American Physical Society meeting in January, 1954. The figure is from M. S. Livingston and J. P. Blewett, Particle Accelerators (New York, 1962), p. 6.


Reseda, 1978, p. 177, Original from
are biased by difficulties in time separation between event and judges, as well as by the judges' (Western) cultural background.

A relevant fact is the relationship of the rate of technology accretion to the rate of productivity increases. The quantity of scientific literature has been doubling at a rate which has been fairly constant for a long time. Price (1961, p. 119) estimates that "normal exponential growth" since 1660 has been a constant doubling every thirty years (whereas before that 120 years were required for doubling). On the other hand, productivity has been increasing at an increasing rate, as discussed by Solow (1957) and Fellner (1970). The latter observation seems inconsistent with the idea that the "quality" of economic literature has been decreasing, or even remaining constant; rather, a constant doubling in the quality of literature together with an increased rate of productivity suggests an increase in economic productivity per unit of scientific literature (taking the latter as a proxy for all knowledge produced).

3. So far we have assumed that the rate of idea production (duplicated plus unduplicated) is the same for representative persons under all conditions. But we know—from statistical as well as casual evidence (Price, 1971; Love and Pashute, 1978)—that the production of technical-advance ideas is a positive function of per-capita income. Building in this effect of income on idea production will only amplify the results shown without it.

Of course each and every fact and line of thought in this section is speculative at best, and irrelevant at worst. But this proves, more than anything, how ill-formed and uninvestigated is this topic. Any
discussion, crude that it may be, may be a useful starting point in such conditions.

As long as we think of technology as scientific knowledge in itself, we are not likely to reach a persuasive answer. But if we keep in mind that our interest here is in the output of economic welfare, we may be able to make some progress. We still will have major difficulties if we think in terms of GNP per capita as it is usually measured, because GNP measurement is itself affected by technological change in a variety of well-known ways. But it may be reasonable to examine the changes over time in a few key elements of economic welfare that seem reasonably comparable.

Consider grain output, for example. Its level of production technology in an economy is related to the level of other production technology, and it is old enough a product so that it is not in a spectacular early-development period. Yet the rate of productivity increase has been increasing, and there seems no reason to doubt that the amount of human labor involved in grain production per capita will continue to diminish in the foreseeable future, and perhaps at an increasing rate (e.g., dropping 1% of its manpower in the first equal period under discussion, 1.1% in the second period, and so on, an infinitely sustainable process until the last farmer). This is at least measurable, and it is a reasonable proxy for the combined quantity and quality of technical knowledge. If the change in the level of productivity of all other products is similar to that of grain, this suggests that a given amount of product can be attained by an amount of work time that will diminish at an increasing rate, while leisure—a key element of economic
welfare—will increase. There is an upper bound to the amount of leisure, of course. But if the amount of leisure is held constant and more types of goods are produced, economic welfare could continue to increase at an increasing rate.*

VII. WHICH TECHNICAL PROGRESS FUNCTION DOES THE MICRO-MODEL OF INVENTION FIT?

We are now in a position to consider which technical-progress models in the literature best fit the micro-model set out above.

Quite obviously, there is no warrant here for technical progress being driven purely exogenously, or being considered simply a function of time. And this model does not fit Arrow's learning-by-doing model, which could be written

\[ (5a) \quad M_t - M_{t-1} = a \left( K^b_t - K^b_{t-1} \right) \]

or by a more straightforward learning-by-doing model such as

\[ (12) \quad M_t - M_{t-1} = a \left[ \sum_{t=0}^{t=\tau} Y_t^b - \left( \sum_{t=0}^{t=\tau-1} Y_t^b \right) \right] \]

The micro-model views technical progress as a function of human activity together with the stock of technology, perhaps in conjunction with the level of education; the learning-by-doing models are purely empirical in their origin, and no technology-producing mechanism is suggested for them. And everything said about learning-by-doing models pertains just as well to the Verdoorn model.

*Life expectancy is an economic good at least as valuable as leisure and GNP goods. And I see no prospect of it increasing at an increasing rate. The reader may take this as a fatal objection to the argument—or may not.
and the Kaldor model

$$\frac{M_t - M_{t-1}}{M_{t-1}} = f\left(\frac{Y_t - Y_{t-1}}{Y_t}\right)$$

Furthermore, the Verdoorn, Arrow, and Kaldor models have different implications than the micro-models as can be seen in Table 1.

The best fit to the micro-model is a transformation of Phelps' (1966) function, which can be written here as

$$M_t - M_{t-1} = f(L_{t-1}^\phi M_{t-1}^\psi), \quad \phi, \psi < 1.$$  

In the limit of the micro-model this becomes

$$M_t - M_{t-1} = f(L_{t-1}^\phi)$$

which may be viewed as equation (7a) with $\psi \to 0$. The Phelps function also fits nicely with a function that brings in the level of education through the income level

$$M_t - M_{t-1} = f[L_{t-1}^\phi M_{t-1}^\psi \frac{Y_{t-1}}{\Delta}]$$

Or if one thinks that the aggregate income of a country should be seen as an additional proxy for the aggregate knowledge available to be worked with, and that this should be brought into the function along with per-capita income, the technical progress function may be written

$$M_t - M_{t-1} = f[L_{t-1}^\phi M_{t-1}^\psi \frac{Y_{t-1} E}{\Delta}]$$

Rather surprisingly, technical progress functions (7c) and (7d) maintain
all the key growth-theoretic features of Phelps' function (7a) until the sum of the exponents L and Y becomes much higher than a function homogeneous of degree one; the characteristics of the models are explored elsewhere (Simon, 1979b). And all of these functions are consistent with the micro-model of invention developed in this paper, at various stages of technical progress. This suggests that functions (7a) - (7d) should be considered the main candidates for inclusion in growth-theoretic and population-policy macro-models.

CONCLUSION

For hundreds of years economists and others have recognized that a larger population and faster population growth seem to be found where individual income is high and rising. But at the same time, the simplest sort of theoretical reasoning suggested exactly the opposite conclusion. It is obvious—Malthus only formalized the notion—that in the shortest run, during which all capital and inventory are fixed, more people imply lower average incomes because there are more mouths to eat up the inventory, and more hands per unit of productive capital. It is, however, almost as obvious—except, perhaps, to some economists—that after some long run, additional people lead to higher per-worker output and per-capita income through increases in productivity due to human invention and adoption, an idea stated by Petty and vigorously restated by Kuznets.

In the past two decades, several lines of empirical evidence have shown increases in productivity due to increased technology, market size, and total output. These include the Abramovitz-Solow-Denison components-of-growth studies, the learning-by-doing studies, the studies of comparative
growth in market size starting with Rostas, and recent studies of productivity by size of city by Sveikauskas, Segal, and Love.

All the existing empirical evidence, however, is at the level of the economy or the industry. Still lacking—aside from Petty's simple suggestion—is an empirical or theoretical understanding at the micro level of why a larger population and a bigger economy should lead to increased productivity. The task of this paper is to propose such a theoretical model. The constituents of the model are the number of potential technology producers, the number of elements available in the environment to stimulate invention, and the probabilistic relationships between inventors and stimuli.

The model works out the tradeoff between the idea-stock-reducing and idea-stock-increasing forces, under different conditions of population growth. The model shows that under assumptions that seem economically and psychologically reasonable, a larger labor force has increasing returns in technology production. The number of possible new combinations that result from the addition of a new idea element to the pool of technical knowledge is very large relative to the depletion of the pre-existing pool of technical knowledge by the discovery of that one idea. And the number of such possible new combinations that a representative additional person creates is very large relative to the additional possibilities of duplicated effort that result from the additional person. Hence the net result of an additional person is an increase in the total number of new ideas. And this effect will continue with increases in total persons and total technology until duplication approaches zero, and unduplicated idea production for the representative person approaches
the person's total capacity for idea production. Hence there are increasing returns in idea-production to additional persons until convergence to constant returns.

As to the quality (or value or importance) of additional ideas with a larger rate of growth of the labor force and higher rates of increase of technology, there is no statistical evidence to suggest that quality is decreasing. On a priori grounds one might speculate, rather, that effective quality is increasing, because of the larger number of existing ideas and capital with which a new idea can interact and be fruitful. But it is logically impossible to evaluate the economic contribution of later ideas separately from the contribution of earlier ideas.

The model also considers the trade-off between an additional person's positive contribution in technology and the person's negative effect on per-worker output due to the Malthusian element of capital dilution. An increase in workers must surely lower individual income in the short run due to capital dilution, but after some time the effect comes to be positive because of the improved technical level. The trade-off between present and future is discussed in another paper (Simon, 1979).
References


