

Endogenous Technological Change

Author(s): Paul M. Romer

Source: *Journal of Political Economy*, Vol. 98, No. 5, Part 2: The Problem of Development: A Conference of the Institute for the Study of Free Enterprise Systems (Oct., 1990), pp. S71-S102

Published by: The University of Chicago Press

Stable URL: <http://www.jstor.org/stable/2937632>

Accessed: 07-08-2017 17:31 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Political Economy*

Endogenous Technological Change

Paul M. Romer

University of Chicago

Growth in this model is driven by technological change that arises from intentional investment decisions made by profit-maximizing agents. The distinguishing feature of the technology as an input is that it is neither a conventional good nor a public good; it is a non-rival, partially excludable good. Because of the nonconvexity introduced by a nonrival good, price-taking competition cannot be supported. Instead, the equilibrium is one with monopolistic competition. The main conclusions are that the stock of human capital determines the rate of growth, that too little human capital is devoted to research in equilibrium, that integration into world markets will increase growth rates, and that having a large population is not sufficient to generate growth.

I. Introduction

Output per hour worked in the United States today is 10 times as valuable as output per hour worked 100 years ago (Maddison 1982). In the 1950s, economists attributed almost all the change in output per hour worked to technological change (Abramovitz 1956; Kendrick 1956; Solow 1957). Subsequent analysis raised our estimates of

Prepared for the conference "The Problem of Economic Development: Exploring Economic Development through Free Enterprise," held at the State University of New York at Buffalo, May 1988. I have benefited from the comments of many seminar and conference participants and two discussants (Rob Vishny, Buffalo, May 1988, and Dale Jorgenson, National Bureau of Economic Research Economic Fluctuations meeting, July 1988). Discussions with Gary Becker, Karl Shell, Robert Lucas, Gene Grossman, and Elhanan Helpman were especially helpful. Research assistance was provided by Danyang Xie. The original work was supported by National Science Foundation grant SES-8618325. It was revised while I was a visitor at the Center for Advanced Study in the Behavioral Sciences and supported by NSF grant BNS87-00864.

[*Journal of Political Economy*, 1990, vol. 98, no. 5, pt. 2]

© 1990 by The University of Chicago. All rights reserved. 0022-3808/90/9805-0015\$01.50

the importance of increases in the effective labor force and the effective stock of capital in generating growth in output per worker (Jorgenson, Gollop, and Fraumeni 1987), but technological change has surely been important as well. The raw materials that we use have not changed, but as a result of trial and error, experimentation, refinement, and scientific investigation, the instructions that we follow for combining raw materials have become vastly more sophisticated. One hundred years ago, all we could do to get visual stimulation from iron oxide was to use it as a pigment. Now we put it on plastic tape and use it to make videocassette recordings.

The argument presented in this paper is based on three premises. The first is that technological change—improvement in the instructions for mixing together raw materials—lies at the heart of economic growth. As a result, the model presented here resembles the Solow (1956) model with technological change. Technological change provides the incentive for continued capital accumulation, and together, capital accumulation and technological change account for much of the increase in output per hour worked.

The second premise is that technological change arises in large part because of intentional actions taken by people who respond to market incentives. Thus the model is one of endogenous rather than exogenous technological change. This does not mean that everyone who contributes to technological change is motivated by market incentives. An academic scientist who is supported by government grants may be totally insulated from them. The premise here is that market incentives nonetheless play an essential role in the process whereby new knowledge is translated into goods with practical value. Our initial understanding of electromagnetism arose from research conducted in academic institutions, but magnetic tape and home videocassette recorders resulted from attempts by private firms to earn a profit.

The third and most fundamental premise is that instructions for working with raw materials are inherently different from other economic goods. Once the cost of creating a new set of instructions has been incurred, the instructions can be used over and over again at no additional cost. Developing new and better instructions is equivalent to incurring a fixed cost. This property is taken to be the defining characteristic of technology.

Most models of aggregate growth, even those with spillovers or external effects, rely on price-taking behavior. But once these three premises are granted, it follows directly that an equilibrium with price taking cannot be supported. Section II of the paper starts by showing why this is so. It also indicates which of the premises is dropped in growth models that do depend on price-taking behavior. The argument in this section is fundamental to the motivation for the particu-

lar model of monopolistic competition that follows, but it is more general than the model itself.

In the specific model outlined in Section III, a firm incurs fixed design or research and development costs when it creates a new good. It recovers those costs by selling the new good for a price that is higher than its constant cost of production. Since there is free entry into this activity, firms earn zero profit in a present value sense.

The conclusions of the model follow directly from this specification. On the basis of results from the static theory of trade with differentiated goods (see, e.g., Helpman and Krugman 1985), one should expect that fixed costs lead to gains from increases in the size of the market and therefore to gains from trade between different countries. Perhaps the most interesting feature of the equilibrium calculated for the model constructed here is that increases in the size of the market have effects not only on the level of income and welfare but also on the rate of growth. Larger markets induce more research and faster growth.

The analysis also suggests why population is not the right measure of market size and why the presence of a large domestic market in countries such as China or India is not a substitute for trade with the rest of the world. The growth rate is increasing in the stock of human capital, but it does not depend on the total size of the labor force or the population. In a limiting case that may be relevant for historical analysis and for the poorest countries today, if the stock of human capital is too low, growth may not take place at all.

These implications of the model are taken up briefly in the final sections of the paper. Section III describes the functional forms that are used to describe the preferences and the technology for the model. It defines an equilibrium that allows for both monopolistic competition and external effects arising from knowledge spillovers. Section IV offers a brief intuitive description of a balanced growth equilibrium for the model. Section V formally characterizes the equilibrium. Section VI describes the welfare properties of the equilibrium. Section VII discusses the connection implied by the model between trade, research, and growth. Algebraic details of the derivations are placed in the Appendix.

II. Rivalry, Excludability, and Nonconvexities

Economists studying public finance have identified two fundamental attributes of any economic good: the degree to which it is rivalrous and the degree to which it is excludable (Cornes and Sandler 1986). Rivalry is a purely technological attribute. A purely rival good has the property that its use by one firm or person precludes its use by an-

other; a purely nonrival good has the property that its use by one firm or person in no way limits its use by another. Excludability is a function of both the technology and the legal system. A good is excludable if the owner can prevent others from using it. A good such as the code for a computer program can be made excludable by means of a legal system that prohibits copying or by means of encryption and copy protection schemes.

Conventional economic goods are both rivalrous and excludable. They are privately provided and can be traded in competitive markets. By definition, public goods are both nonrival and nonexcludable. Because they are nonexcludable, they cannot be privately provided or traded in markets. Public goods can be introduced into a model of price-taking behavior by assuming the existence of a government that can levy taxes. Basic scientific research is an example of a public good that could be provided in this way and that is relevant for modeling growth.

Rivalry and excludability are closely linked because most rival goods are excludable. (A parking space in a shopping center parking lot is an example of a good that is effectively nonexcludable because the cost of enforcing excludability is too high relative to the value of the good.) The interesting case for growth theory is the set of goods that are nonrival yet excludable. The third premise cited in the Introduction implies that technology is a nonrival input. The second premise implies that technological change takes place because of the actions of self-interested individuals, so improvements in the technology must confer benefits that are at least partially excludable. The first premise therefore implies that growth is driven fundamentally by the accumulation of a partially excludable, nonrival input.

To evaluate these claims, it helps to have a specific case in mind. The example of a nonrival input used in what follows is a design for a new good. The vast majority of designs result from the research and development activities of private, profit-maximizing firms. A design is, nonetheless, nonrival. Once the design is created, it can be used as often as desired, in as many productive activities as desired.

In this sense, a design differs in a crucial way from a piece of human capital such as the ability to add. The design is nonrival but the ability to add is not. The difference arises because the ability to add is inherently tied to a physical object (a human body) whereas the design is not.¹ The ability to add is rivalrous because the person who

¹ The original version of this paper used the terms "embodied" and "disembodied" to refer to the difference between an intangible such as the ability to add, which is tied to a specific person, and an intangible such as a design, which is not. This choice of terminology is not used in this revision because embodiment has another meaning in growth theory and because the notion of rivalry already exists in the public finance literature.

possesses this ability cannot be in more than one place at the same time; nor can this person solve many problems at once. As noted above, rivalry leads to a presumption that human capital is also excludable. Thus human capital can be privately provided and traded in competitive markets. In contrast, the design is nonrival because it is independent of any physical object. It can be copied and used in as many different activities as desired.

Like any scientific concept, nonrivalry is an idealization. It is sometimes observed that a design cannot be a nonrival good because it is itself tied to the physical piece of paper or the physical computer disk on which it is stored. What is unambiguously true about a design is that the cost of replicating it with a drafter, a photocopier, or a disk drive is trivial compared to the cost of creating the design in the first place. This is not true of the ability to add. Training the second person to add is as costly as training the first. For simplicity, the arguments here will treat designs as idealized goods that are not tied to any physical good and can be costlessly replicated, but nothing hinges on whether this is literally true or merely close to being true.

Nonrivalry has two important implications for the theory of growth. First, nonrival goods can be accumulated without bound on a per capita basis, whereas a piece of human capital such as the ability to add cannot. Each person has only a finite number of years that can be spent acquiring skills. When this person dies, the skills are lost, but any nonrival good that this person produces—a scientific law; a principle of mechanical, electrical, or chemical engineering; a mathematical result; software; a patent; a mechanical drawing; or a blueprint—lives on after the person is gone. Second, treating knowledge as a nonrival good makes it possible to talk sensibly about knowledge spillovers, that is, incomplete excludability. These two features of knowledge—unbounded growth and incomplete appropriability—are features that are generally recognized as being relevant for the theory of growth. What thinking about nonrivalry shows is that these features are inextricably linked to nonconvexities.

If a nonrival input has productive value, then output cannot be a constant-returns-to-scale function of all its inputs taken together. The standard replication argument used to justify homogeneity of degree one does not apply because it is not necessary to replicate nonrival inputs. Suppose that a firm can invest 10,000 hours of engineering time to produce a design for a 20-megabyte hard disk drive for computers. Suppose that it can produce a total of 2 trillion megabytes of storage per year (i.e., 100,000 units of the drive) if it builds a \$10 million factory and hires 100 workers. If it merely replicates the rival inputs—the factory and the workers—it can double its output to 4 trillion megabytes of storage per year.

Suppose that the firm could have invested 20,000 hours of en-

gineering time in the design work instead of 10,000 hours and, by doing so, could have designed a 30-megabyte hard disk drive that could be manufactured with the same factory and workers. When the firm doubles all its inputs, it uses a 20,000-hour design, two factories, and 200 workers and produces 6 trillion megabytes of storage per year, three times the original output.

More formally, if $F(A, X)$ represents a production process that depends on rival inputs X and nonrival inputs A , then by a replication argument, it follows that $F(A, \lambda X) = \lambda F(A, X)$. This replication argument assumes that X is an exhaustive list of the rival inputs. Because the focus here is on national economies, the argument neglects integer problems that may be relevant for a small market that gets stuck between n and $n + 1$ plants. The fact that it may not be possible to actually replicate all the inputs in the list X has no bearing on this argument about the properties of $F(\cdot)$.

If A is productive as well, it follows that F cannot be a concave production function because $F(\lambda A, \lambda X) > \lambda F(A, X)$. Because of the properties of homogeneous functions, it also follows that a firm with these kinds of production possibilities could not survive as a price taker. If disk drives sold for marginal cost, annual revenue for the firm would just equal interest payments on the capital and wage payments to workers. More generally, since $F(A, X) = X \cdot (\partial F / \partial X)(A, X)$, it follows that

$$F(A, X) < A \cdot \frac{\partial F}{\partial A}(A, X) + X \cdot \frac{\partial F}{\partial X}(A, X).$$

If all inputs were paid their value marginal product, the firm would suffer losses.

This point has been made many times before (Schumpeter 1942; Arrow 1962*b*; Shell 1966, 1967, 1973; Nordhaus 1969; Wilson 1975). Previous growth models have avoided this difficulty in various ways. Solow (1956) treats A as an exogenously provided public input (i.e., an input that is both nonexcludable and nonrival). Shell (1966, 1967) treats it as a public input that is provided by the government. In each case, the factor A receives no compensation, and every individual firm is assumed to be free to exploit the entire stock of A . These models are consistent with the first premise, that technological change drives growth, and the third, that the technology is a nonrival good, but they are inconsistent with the second premise. They both deny the role that private, maximizing behavior plays in generating technological change.

In an attempt to make the evolution of A responsive to market incentives, Arrow (1962*a*) assumed that an increase in K necessarily leads to an equiproportionate increase in knowledge through "learn-

ing by doing," but he still treats knowledge as a public good. Lucas (1988) assumed in effect that it is production of human capital rather than physical capital that generates this nonrival, nonexcludable good. Both of these papers make the production of a nonrival, nonexcludable good an unintentional side effect of the production of a conventional good.

The learning-by-doing formulation has the advantage that it makes the rate of accumulation of nonrival knowledge endogenous, but it is unsatisfactory because it takes the strict proportionality between knowledge and physical capital or knowledge and education as an unexplained and exogenously given feature of the technology. It preserves the public-good character of knowledge assumed by Solow and Shell but makes it a public good that is privately provided as a side effect. Like the other public-good formulations, it rules out the possibility that firms make intentional investments in research and development.

This formulation has the additional difficulty that it is not robust. The nonrival input produced through learning by doing must be completely nonexcludable. If it were even partially excludable, Dasgupta and Stiglitz (1988) show that decentralized equilibrium with many firms would not be sustainable.

In a partial equilibrium model of an industry in which firms face upward-sloping cost curves, Shell (1973) proposed a model with price taking in which expenditure on research was compensated out of quasi rents. Griliches (1979), again in an industry setting, made this formulation more explicit. He assumed that the production function takes the form $F(A_N, A_E, X)$, where A_E represents an excludable part of the benefits of research and development and A_N represents the nonexcludable part. Since A_E is excludable, it is accumulated intentionally. The nonexcludable part A_N is created as a side effect of producing A_E . He also assumed that the function $F(\cdot)$ is homogeneous of degree one in X and A_E taken together.

In an aggregate model of growth, I made the same kind of assumption (Romer 1986). To make the dynamic analysis in this paper simple, I reduced the dynamic model to one with a single-state-variable model by assuming that the excludable good A_E that the firm produces intentionally is used in fixed proportions with physical capital. As a result, the model ends up having dynamics similar to those of Arrow's learning-by-doing model, and the mathematical equations can be interpreted equally well in terms of learning by doing that is incidental to capital production.

The advantage of the interpretation that knowledge is compensated out of quasi rents is that it allows for intentional private investments in research and development. The difficulty is that it violates

the logic of the replication argument. If the input A_E is truly the result of research and development, it is a nonrival good. In this case, the function $F(\cdot)$ must be homogeneous of degree one in X alone and the Dasgupta and Stiglitz (1988) argument applies. If a nonrival input is even partially excludable, nonconvexities are present and a decentralized equilibrium cannot be sustained. Stated in terms of cost curves, the replication argument implies that long-run cost curves are horizontal when all rival inputs are treated as variable. What appear to be quasi rents are merely competitive returns to rival factors that are in fixed supply. These quasi rents cannot be used to compensate both the innovation activity and the rival fixed factors. Even if these factors are in fixed supply at the aggregate level, they presumably have alternative uses and will not be supplied to an activity if they are not paid their marginal product.

An alternative approach to growth theory with price-taking competition is to dispute the first and third premises, that technological change drives growth and that knowledge about the technology is a nonrival input. Human capital models such as those presented by King and Rebelo (1987), Jones and Manuelli (1988), Rebelo (1988), and Becker, Murphy, and Tamura (this issue) treat all forms of intangible knowledge as being analogous to human capital skills that are rivalrous and excludable. There is no nonrival input like the technology and, hence, no nonconvexities or spillovers.

The only way to accept all three premises described in the Introduction is to return to the suggestion of Schumpeter (1942) and explicitly introduce market power. Shell (1973) described a model with a single monopolist who invests in technological change, but as he recognizes, it is difficult to give an aggregate interpretation to a model dominated by a single firm. In a recent paper (Romer 1987), I presented a model with market power but also with free entry and many firms. It builds on the model of monopolistic competition in consumption goods formulated by Dixit and Stiglitz (1977), applied in a dynamic setting by Judd (1985), used in a dynamic model of trade by Grossman and Helpman (1989*c*), and extended to differentiated inputs in production by Ethier (1982). The specification used here differs from that in my earlier paper primarily because it emphasizes the importance of human capital in the research process. The earlier model showed that scale is an important determinant of the rate of growth. The analysis here shows that the correct measure of scale is not population but human capital.

III. Description of the Model

The four basic inputs in this model are capital, labor, human capital, and an index of the level of the technology. Capital is measured in

units of consumption goods. Labor services L are skills such as eye-hand coordination that are available from a healthy physical body. They are measured by counts of people. As used here, human capital H is a distinct measure of the cumulative effect of activities such as formal education and on-the-job training.

The concept of human capital as years of education or training that are person specific is close to the one used in labor market contexts (e.g., Heckman 1976; Rosen 1976). It corresponds to the practice in growth accounting applications that take account of changes in the quality of the labor force due to changes in observables such as the level of education and experience (see, e.g., Gollop and Jorgenson 1980). This concept of human capital is more limited than the notion used in theoretical models of growth based on unlimited human capital accumulation such as those presented by King and Rebelo (1987), Lucas (1988), and Becker et al. (this issue). These models implicitly combine a notion of knowledge that can outlive any individual with a labor market notion of human capital that does not.

The model used here separates the rival component of knowledge, H , from the nonrival, technological component, A . Because it has an existence that is separate from that of any individual, A can grow without bound. In the specific formulation used below, each new unit of knowledge corresponds to a design for a new good, so there is no conceptual problem measuring A . It is a count of the number of designs.

The formal model of the economy has three sectors. The research sector uses human capital and the existing stock of knowledge to produce new knowledge. Specifically, it produces designs for new producer durables. An intermediate-goods sector uses the designs from the research sector together with forgone output to produce the large number of producer durables that are available for use in final-goods production at any time. In practice, one might expect research on a new design and the production of the new good to take place within the same firm, and nothing in the analysis here rules this out; design work can take place either internally or in a separate firm that sells its patent to the firm that will produce the actual good. A final-goods sector uses labor, human capital, and the set of producer durables that are available to produce final output. Output can be either consumed or saved as new capital.

To keep the dynamic analysis simple and highlight the effects of interest, several simplifying assumptions are used. The first is that the population and the supply of labor are both constant. This rules out an analysis of fertility, labor force participation, or variation in hours worked per worker. The second is that the total stock of human capital in the population is fixed and that the fraction supplied to the market is also fixed. Thus the supply of the aggregate factors L and H

is fixed. The assumption on H is made largely for technical reasons. The dynamic analysis is greatly simplified by restricting attention to equilibria with constant growth rates. In a stationary population in which people have finite lives, the only feasible constant growth rate for total years of education or experience is zero. A more complicated dynamic analysis could consider the effects of the kinds of increases in H and L that have been observed historically, but even in this kind of analysis, H must ultimately approach an upper bound.

The other simplifying assumptions are extreme assumptions on factor intensities. One has already been made implicitly. Assuming that capital can be accumulated as forgone output is equivalent to assuming that capital goods are produced in a separate sector that has the same technology as the final-output sector. Forgoing consumption is then equivalent to shifting resources from the consumption sector into the capital sector. Also, the plausible assertion that research is relatively human capital- and knowledge-intensive is translated into an extreme specification in which only knowledge and human capital are used to produce new designs or knowledge. Labor and capital do not enter at all. These kinds of restrictions will reduce the analysis of the dynamics of this system to a system of equations that can be explicitly solved by doing algebra. Presumably, a relaxation of these assumptions that preserves the factor intensity orderings used here would not change the basic dynamics of the model.

Final output Y in this model is expressed as a function of physical labor L , human capital devoted to final output H_Y , and physical capital. The unusual feature of the production technology assumed here is that it disaggregates capital into an infinite number of distinct types of producer durables. For now, let these durables be indexed by an integer i . (Soon, the index i will be assumed to be a continuous variable instead of a discrete one to avoid integer constraints.) Only a finite number of these potential inputs, the ones that have already been invented and designed, are available for use at any time. Thus if $x = \{x_i\}_{i=1}^{\infty}$ is the list of inputs used by a firm that produces final output, there is some value A such that $x_i = 0$ for all $i \geq A$. Because A changes as new producer durables are invented, it is important to be able to describe final output as a stationary function of all conceivable input lists.

In this kind of environment, a simple functional form for output is the following extension of the Cobb-Douglas production function:

$$Y(H_Y, L, x) = H_Y^{\alpha} L^{\beta} \sum_{i=1}^{\infty} x_i^{1-\alpha-\beta}. \quad (1)$$

This production function differs from the usual production function

only in its assumption about the degree to which different types of capital goods are substitutes for each other. In the conventional specification, total capital K is implicitly defined as being proportional to the sum of all the different types of capital. This definition implies that all capital goods are perfect substitutes. One additional dollar of capital in the form of a truck has the same effect on the marginal productivity of mainframe computers as an additional dollar's worth of computers. Equation (1) expresses output as an additively separable function of all the different types of capital goods so that one additional dollar of trucks has no effect on the marginal productivity of computers.

Among the various types of capital goods, one can imagine pairs of inputs that are close substitutes (trucks and trains), pairs that are complements (computers and communications networks), and many pairs that fall somewhere in between. The conventional formulation explores the case in which all durables are perfect substitutes. The model here considers the case in which all durables have additively separable effects on output. An investigation of complementarity as well as of mixtures of types of substitutability is left for future work.

Because the production function in equation (1) is homogeneous of degree one, output in the final-goods sector can be described in terms of the actions of a single, aggregate, price-taking firm. The sector that produces producer durables, however, cannot be described by a representative firm. There is a distinct firm i for each durable good i . A firm must purchase or produce a design for good i before commencing production. Once it owns the design, the firm can convert η units of final output into one durable unit of good i . As in the standard one-sector model, the formal specification here describes the sector that produces capital goods as a black box that takes final output in on one side and gives capital goods out of the other side. The correct interpretation of this formal description is that the forgone consumption is never manufactured. The resources that would have been used to produce the forgone output are used instead to manufacture capital goods. It is possible to exchange a constant number of consumption goods for each unit of capital goods if the production function used to manufacture capital goods has exactly the same functional form as the production function used to manufacture consumption goods.

Once a firm has produced a design for durable i , it can obtain an infinitely lived patent on that design. If the firm manufactures $x(i)$ units of the durable, it rents those durables to final-output firms for a rental rate $p(i)$. Since firm i will be the only seller of capital good i , it will face a downward-sloping demand curve for its good. Since the durables are assumed not to depreciate, the value of one unit of durable i is the

present discounted value of the infinite stream of rental income that it generates.

There are many equivalent institutional arrangements that can support any given equilibrium. In what follows, the firm that owns the patent on the design for durable i is assumed to be the only one that manufactures it, but this is only a convenience. Whether the owner of the patent manufactures the good itself or licenses others to do so, it can extract the same monopoly profit. Design of new durables and manufacturing could take place within the same firm, but it is easier to describe the equilibrium if the research and development department is treated as a separate firm and designs are transferred for an explicit price. It is also easier to assume that the firm that buys a design and manufactures a differentiated producer durable rents its durables instead of selling them outright. In particular, this shows that there are market mechanisms that avoid the usual durable-goods-monopoly problem. Provided that the manufacturer of the durable could commit to levels of output, nothing would change if the durables were sold instead of rented. The analysis is further simplified by assuming that the durables do not depreciate. Adding depreciation would merely add a familiar term to the user cost of capital.

In parallel with the usual one-sector model and in conformity with national income accounting conventions, it is useful to define an accounting measure of total capital K as cumulative forgone output. Thus $K(t)$ evolves according to the rule

$$\dot{K}(t) = Y(t) - C(t), \quad (2)$$

where $C(t)$ denotes aggregate consumption at time t . Because it takes η units of forgone consumption to create one unit of any type of durable, this accounting measure K is related to the durable goods that are actually used in production by the rule $K = \eta \sum_{i=1}^{\infty} x_i = \eta \sum_{i=1}^A x_i$.

Thus H and L are fixed, and K grows by the amount of forgone consumption. It remains to specify the process for the accumulation of new designs, that is, for the growth of $A(t)$. As noted above, research output depends on the amount of human capital devoted to research. It also depends on the stock of knowledge available to a person doing research. If designs were treated as discrete indivisible objects that are not produced by a deterministic production process, the production technology for designs would have to take explicit account of both integer constraints and uncertainty. There is no doubt that both indivisibility and uncertainty are important at the micro level and over short periods of time. The simplifying assumption made here is that neither is crucial to a first-pass analysis of technological change at the aggregate level. Henceforth, the index i

for the different types of goods is treated as a continuous variable, and the sum in equation (1) is replaced by an integral:

$$Y(H_Y, L, x) = H_Y^\alpha L^\beta \int_0^\infty x(i)^{1-\alpha-\beta} di. \quad (1')$$

(Here x must be interpreted as a function from an appropriately defined function space.) With this formal structure, the output of new designs produced by researcher j can be written as a continuous, deterministic function of the inputs applied. If the researcher possesses an amount of human capital H^j and has access to a portion A^j of the total stock of knowledge implicit in previous designs, the rate of production of new designs by researcher j will be $\delta H^j A^j$, where δ is a productivity parameter.

This formulation with a continuum of goods is close to that used by Judd (1985) in his discussion of patents. He also studies the process whereby goods are introduced as new patents are produced. The model here corresponds to Judd's discussion of the case in which patents are infinitely lived. The main differences are that Judd treats the differentiated goods as consumption goods rather than producer durables and uses a form of exogenous technological change to generate growth of productivity in the research sector. Here, growth in A by itself increases the productivity of human capital in the research sector.

Although other assumptions about secrecy and property rights could be considered, the equilibrium here is based on the assumption that anyone engaged in research has free access to the entire stock of knowledge. This is feasible because knowledge is a nonrival input. All researchers can take advantage of A at the same time. The output of researcher j is therefore $\delta H^j A$. If we sum across all people engaged in research, the aggregate stock of designs evolves according to

$$\dot{A} = \delta H_A A, \quad (3)$$

where H_A has the obvious interpretation of total human capital employed in research.

Equation (3) contains two substantive assumptions and two functional form assumptions. The first substantive assumption is that devoting more human capital to research leads to a higher rate of production of new designs. The second is that the larger the total stock of designs and knowledge is, the higher the productivity of an engineer working in the research sector will be. According to this specification, a college-educated engineer working today and one working 100 years ago have the same human capital, which is measured in terms of years of forgone participation in the labor market. The engineer working today is more productive because he or she can take advan-

tage of all the additional knowledge accumulated as design problems were solved during the last 100 years.

The two functional form assumptions are that the output of designs is linear in each of H_A and A when the other is held constant. These assumptions, like the exclusion of the inputs L and $x(\cdot)$, are made largely for analytical convenience. Linearity in H_A is not important for the dynamic properties of the model, but weakening this assumption would require a more detailed specification of how income in the research sector is allocated to the participants.

Linearity in A is what makes unbounded growth possible, and in this sense, unbounded growth is more like an assumption than a result of the model. In what follows, it will become clear that the marginal product of human capital H_Y employed in the manufacturing sector grows in proportion to A . If A were replaced in equation (3) by some concave function of A —that is, if the marginal productivity of human capital in the research sector does not continue to grow in proportion to A —human capital employed in research would shift out of research and into manufacturing as A becomes larger. This will cause the rate of growth to slow down. Whether opportunities in research are actually petering out, or will eventually do so, is an empirical question that this kind of theory cannot resolve. The specification here, in which unbounded growth at a constant rate is feasible, was chosen because there is no evidence from recent history to support the belief that opportunities for research are diminishing. Moreover, linearity in A is convenient analytically, and assumptions about what will happen in the far future for values of A that are very large relative to the current level have very little effect on the question of interest: How do other variables in the model affect the rate of growth of A ?

The crucial feature of the specification used here is that knowledge enters into production in two distinct ways. A new design enables the production of a new good that can be used to produce output. A new design also increases the total stock of knowledge and thereby increases the productivity of human capital in the research sector. The owner of a design has property rights over its use in the production of a new producer durable but not over its use in research. If an inventor has a patented design for widgets, no one can make or sell widgets without the agreement of the inventor. On the other hand, other inventors are free to spend time studying the patent application for the widget and learn knowledge that helps in the design of a widget. The inventor of the widget has no ability to stop the inventor of a widget from learning from the design of a widget. This means that the benefits from the first productive role for a design are completely

excludable, whereas the benefits from the second are completely non-excludable. In an overall sense, this means that the nonrival design inputs are partially excludable.

In theory and in practice, there is always some ambiguity about what constitutes a design for a new and different good and what constitutes a copy of an existing design. In the model, this ambiguity is artificially resolved by the form of the production function Y . This functional form implies that new goods are never close substitutes for existing goods since all the producer durables enter into production in an additively separable fashion. (See Pascoa [1986] for a discussion of this general property of "no neighboring goods.") Although it greatly simplifies the analysis, this is not a realistic feature of the model. In particular, it rules out the possibility of obsolescence. Nonetheless, the general results here should be robust to more careful modeling of the nature of the interaction between different specialized producer durables. What matters for the results is that the knowledge is a nonrival good that is partially excludable and privately provided.

At the aggregate level, H_A and H_Y are related by the constraint $H_Y + H_A = H$. According to these equations, any person can devote human capital to either the final-output sector or the research sector. Implicitly this formulation neglects the fact that L and H are supplied jointly. To take the equations used here literally, one must imagine that there are some skilled persons who specialize in human capital accumulation and supply no labor.

To fix notation for prices, let spot prices at any point in time be measured in units of current output and let r denote the interest rate on loans denominated in goods. Let P_A denote the price of new designs, and let w_H denote the rental rate per unit of human capital. Because goods can be converted into capital one for one, the spot price for capital is one and its rate of return is r . Because of the assumption that anyone engaged in research can freely take advantage of the entire existing stock of designs in doing research to produce new designs, it follows from equation (3) that P_A and w_H are related by $w_H = P_A \delta A$.

Once a design has been produced, a large number of potential suppliers of the new good bid for the right to do so. Each of these firms takes the price P_A for designs, the price of one for capital goods, and the interest rate as given, but if it begins production it sets prices to maximize profits. Formally, it helps to let the rental price $p(i)$ for the i th durable lie in the range $\mathbb{R}_+ \cup \{\infty\}$ for its durable good. If no firm produces good i , its price can be understood to be $p(i) = \infty$.

Faced with a price list $\{p(i): i \in \mathbb{R}_+\}$ for all the producer durables,

including infinite prices for the durables that have not been invented yet, the representative final-output firm chooses a profit-maximizing quantity $x(i)$ for each durable. Because it is a constant-returns-to-scale firm, its input demands are defined only after the scale of operation is pinned down. Let L and H_Y be the total amounts of labor and human capital that are used in the production of final-output goods. (The split of total H between H_Y and H_A remains to be determined.) Given values for H_Y and L , it is possible to derive the aggregate demand for the durables from a maximization problem that is conditional on them:

$$\max_x \int_0^\infty [H_Y^\alpha L^\beta x(i)^{1-\alpha-\beta} - p(i)x(i)] di.$$

Differentiating under the integral sign leads to an inverse demand function

$$p(i) = (1 - \alpha - \beta)H_Y^\alpha L^\beta x(i)^{-\alpha-\beta}. \quad (4)$$

(There is an important technical issue about what it means to the final-goods producer if prices change on a set of values of i that has measure zero. Because of the symmetry in this model, eq. [4] can readily be derived by a limiting argument. For a general discussion of this issue, see Pascoa [1986, 1990].)

The demand curve in equation (4) is what the producer of each specialized durable takes as given in choosing the profit-maximizing price to set. Faced with given values of H_Y , L , and r , a firm that has already incurred the fixed-cost investment in a design will choose a level of output x to maximize its revenue minus variable cost at every date:

$$\begin{aligned} \pi &= \max_x p(x)x - r\eta x \\ &= \max_x (1 - \alpha - \beta)H_Y^\alpha L^\beta x^{1-\alpha-\beta} - r\eta x. \end{aligned} \quad (5)$$

The flow of rental income is $p(x)$ times x . The cost is the interest cost on the ηx units of output needed to produce x durables. To keep the analysis simple, it is convenient to assume that the capital is putty-putty, so that the firm can solve this problem at every point in time, converting units of durables back into general capital, and avoid the interest cost if it decides to supply fewer units. At any date, the only sunk cost is the initial expenditure on the design. This assumption is harmless because the demand for durables is stationary in equilibrium, so no disinvestment ever takes place.

The monopoly pricing problem specified in equation (5) is that of a

firm with constant marginal cost that faces a constant elasticity demand curve. The resulting monopoly price is a simple markup over marginal cost, where the markup is determined by the elasticity of demand, $\bar{p} = r\eta/(1 - \alpha - \beta)$. The flow of monopoly profit is $\pi = (\alpha + \beta)\bar{p}\bar{x}$, where $\bar{x}$ is the quantity on the demand curve (4) implied by the price $\bar{p}$.

Each producer of specialized durables must rent its output to a large number of final-goods producers that can operate at any scale. By assumption, it is not possible for the producer to monitor the use of its durables. As a result, price discrimination is not feasible. The best the firm can do is charge the simple monopoly price.

The decision to produce a new specialized input depends on a comparison of the discounted stream of net revenue and the cost P_A of the initial investment in a design. Because the market for designs is competitive, the price for designs will be bid up until it is equal to the present value of the net revenue that a monopolist can extract. At every date t , it must therefore be true that

$$\int_t^\infty e^{-\int_t^s r(s)ds} \pi(\tau) d\tau = P_A(t). \quad (6)$$

If P_A is constant (as it will be in the equilibrium described below), this condition can be put in a more intuitive form. Differentiating with respect to time t yields

$$\pi(t) - r(t) \int_t^\infty e^{-\int_t^s r(s)ds} \pi(\tau) d\tau = 0.$$

Substituting in the expression for P_A from equation (6) yields

$$\pi(t) = r(t)P_A. \quad (6')$$

This equation says that at every point in time, the instantaneous excess of revenue over marginal cost must be just sufficient to cover the interest cost on the initial investment in a design. (This formulation of the intertemporal zero profit constraint is taken from Grossman and Helpman [1989c].)

The solution of the model for a balanced growth equilibrium given in Section V shows that the technology described above implies a negatively sloped linear relation between the rate of growth of output and the rate of return on investment. To close the model, it remains to specify preferences that imply a parallel relation between the rate of growth of consumption and the marginal rate of intertemporal substitution. This relation is easily derived for Ramsey consumers with discounted, constant elasticity preferences:

$$\int_0^{\infty} U(C)e^{-\rho t} dt, \quad \text{with } U(C) = \frac{C^{1-\sigma} - 1}{1-\sigma} \quad \text{for } \sigma \in [0, \infty).$$

The implied intertemporal optimization condition for a consumer faced with a fixed interest rate r is that $\dot{C}/C = (r - \rho)/\sigma$. Preferences enter the solution of the model only through this relation between the consumption growth rate and the interest rate.

The consumers are endowed with fixed quantities of labor L and human capital H that are supplied inelastically. At time 0, consumers own the existing durable-goods-producing firms, and the net revenues of these firms are paid to consumers as dividends. Final-goods firms earn zero profits and own no assets, so they can be ignored in the specification of endowments.

An equilibrium for this model will be paths for prices and quantities such that (i) consumers make savings and consumption decisions taking interest rates as given; (ii) holders of human capital decide whether to work in the research sector or the manufacturing sector taking as given the stock of total knowledge A , the price of designs P_A , and the wage rate in the manufacturing sector w_A ; (iii) final-goods producers choose labor, human capital, and a list of differentiated durables taking prices as given; (iv) each firm that owns a design and manufactures a producer durable maximizes profit taking as given the interest rate and the downward-sloping demand curve it faces, and setting prices to maximize profits; (v) firms contemplating entry into the business of producing a durable take prices for designs as given; and (vi) the supply of each good is equal to the demand.

IV. Discussion of the Model

A reasonable intuition for the behavior of this model can be inferred by considering the Solow (1956) model, in which the evolution of A is given exogenously, and the Uzawa (1965) model, in which the evolution of A is determined by the allocation of resources between a research sector and a final-goods sector. For a fixed amount of A , and therefore a fixed set of producer durables, the model is almost identical to the Solow model. Because of the symmetry in the model, all the durable goods that are available are supplied at the same level, henceforth denoted as $\bar{x}$. If they were not, it would be possible to increase profits in the producer durable sector by reducing the output of high-output firms and diverting the capital released in this way to low-output goods. Since A determines the range of durables that can be produced and since η units of capital are required per unit of durable

goods, it is possible to solve for $\bar{x}$ from the equation $K = \eta A \bar{x}$. Then output Y can be written as

$$\begin{aligned}
 Y(H_A, L, x) &= H_Y^\alpha L^\beta \int_0^\infty x(i)^{1-\alpha-\beta} di \\
 &= H_Y^\alpha L^\beta A \bar{x}^{1-\alpha-\beta} \\
 &= H_Y^\alpha L^\beta A \left(\frac{K}{\eta A} \right)^{1-\alpha-\beta} \\
 &= (H_Y A)^\alpha (L A)^\beta (K)^{1-\alpha-\beta} \eta^{\alpha+\beta-1}.
 \end{aligned} \tag{7}$$

The last line of this equation shows that the model behaves just like the neoclassical model with labor and human capital augmenting technological change. In particular, it exhibits the usual diminishing returns to capital accumulation. Given the assumed form of preferences, a fixed level of A will lead to an equilibrium with a steady state in which the level of K is determined by the requirement that the marginal product of capital is equal to the discount rate. If A grew at an exogenously specified exponential rate, the economy would converge to a path on which K grows at the same exponential rate as A , just as it does in the Solow model. Along the transition path, the ratio of K to A would change, which implies that r and $\bar{x}$ would change as well. Along the balanced growth path, r , $\bar{x}$, and the ratio of K to A are all constant.

The nonconvexity evident in the expression for final output as a function of the primary inputs of the model (H , L , K , and A) is supported in a decentralized equilibrium that relies on monopolistic competition. In contrast, the nonconvexity present in equation (3) describing output of designs is supported through competition with external effects that arise from knowledge spillovers. In each case, the nonconvexity arises because the nonrival good A is an input in production. In the final-output sector, A matters indirectly because of its effects on the availability of the new $x(\cdot)$ goods. In the research sector, A enters directly.

Both spillovers and price setting seem essential to capturing the features of knowledge in a model of growth. There is little doubt that much of the value to society of any given innovation or discovery is not captured by the inventor, and any model that missed these spillovers would miss important elements of the growth process. Yet it is still the case that private, profit-maximizing agents make investments in the creation of new knowledge and that they earn a return on these investments by charging a price for the resulting goods that is greater than the marginal cost of producing the goods.

V. Solution of the Model for a Balanced Growth Equilibrium

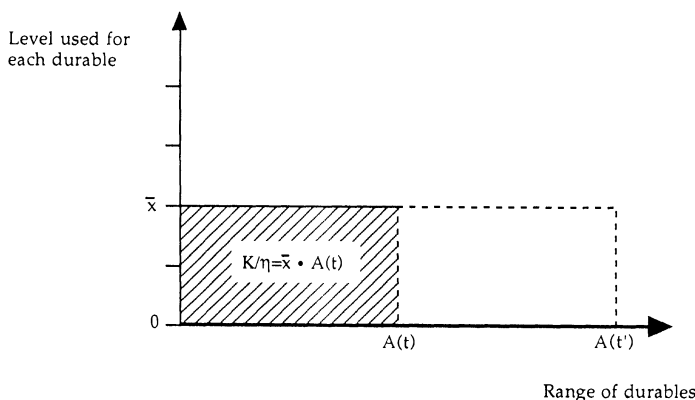
The strategy for characterizing the model that is followed here is to solve for an equilibrium in which the variables A , K , and Y grow at constant exponential rates. This is generally referred to as a balanced growth equilibrium. The intuition from the Solow model suggests that such an equilibrium will exist if A grows at a constant exponential rate. The intuition from the Uzawa model suggests that it is possible for A to grow at an exponential rate because equation (3) for $\dot{A}$ is linear in A . It will grow at a constant rate if the amount of human capital H_A that is devoted to research stays constant. Verifying that a balanced growth equilibrium exists therefore reduces to the problem of showing that prices and wages are such that H_Y and H_A remain constant as Y , K , C , and A grow.

By focusing only on balanced growth paths, the analysis neglects the transient dynamics that arise when the economy starts from a ratio of K to A that differs from the ratio that is maintained along the balanced growth path. One should be able to study convergence to the balanced growth ratio of K to A using the tools used for studying the Solow and Uzawa models, but this analysis is not attempted here.

The first step in the solution of the model is to derive the relation between the growth rate of output and the rate of return on investment. Coupled with the relation between interest rates and growth rates implied by the preference side of the model, this will determine the rate of growth and interest rate. Given the intuition from the Solow model and the results derived in my earlier model of differentiated inputs (Romer 1987), it follows that along the balanced growth path, the ratio of K to A should be constant, which implies that $\bar{x}$ is constant as well. Because of the accumulation of both K and A , the wage paid for human capital in the final-output sector will grow in proportion to A , but by equation (3), the productivity of human capital in research also grows in proportion to A . Since the productivity of human capital grows at the same rate in both sectors, H_Y and H_A will remain constant if the price P_A for new designs is constant.

Figure 1 illustrates the behavior of the inputs in the model. All the producer durables that have been designed up to time t are used at the level $\bar{x}$ that indicates the height of the rectangle. The width is the measure of the number of designs or durables in use, $A(t)$. The area $A(t)\bar{x}$ is equal to total capital divided by η . Over time, $\bar{x}$ remains constant and A grows at a constant exponential rate.

It remains to check that this description of a balanced growth path is consistent with all the equilibrium conditions. As noted in the dis-

FIG. 1.—Producer durables used in production at time t and $t' > t$

cussion following the monopoly profit problem (5), the flow of profit that can be extracted by the seller of any particular durable input is $\pi = (\alpha + \beta)\bar{p}\bar{x}$. Since the present discounted value of this stream of profit must equal the price P_A of the design, it follows that

$$P_A = \frac{1}{r} \pi = \frac{\alpha + \beta}{r} \bar{p}\bar{x} = \frac{\alpha + \beta}{r} (1 - \alpha - \beta) H_Y^\alpha L^\beta \bar{x}^{1-\alpha-\beta}. \quad (8)$$

The last equality in this expression follows by using equation (4) to evaluate $\bar{p}$ in terms of $\bar{x}$.

The condition determining the allocation of human capital between the final-output and research sectors says that the wages paid to human capital in each sector must be the same. In the final-output sector, the wage for human capital is its marginal product. Since human capital receives all the income from the research sector, the wage there is $P_A \delta A$. To equalize returns to human capital in both sectors, $H_Y = H - H_A$ must be chosen so that

$$\begin{aligned} w_H &= P_A \delta A = \alpha H_Y^{\alpha-1} L^\beta \int_0^\infty \bar{x}^{1-\alpha-\beta} di \\ &= \alpha H_Y^{\alpha-1} L^\beta A \bar{x}^{1-\alpha-\beta}. \end{aligned} \quad (9)$$

Substituting P_A from equation (8) into equation (9) and simplifying yield

$$H_Y = \frac{1}{\delta} \frac{\alpha}{(1 - \alpha - \beta)(\alpha + \beta)} r. \quad (10)$$

For a fixed value of $H_A = H - H_Y$, the implied exponential growth rate for A is δH_A . From the monopoly pricing problem, we know that

$\bar{x}$ is constant if r is. The intuition from the Solow model suggests, and examination of the expression for final output

$$Y = H_Y^\alpha L^\beta \int_0^\infty \bar{x}^{1-\alpha-\beta} di = H_Y^\alpha L^\beta A \bar{x}^{1-\alpha-\beta}$$

from equation (7) shows, that output grows at the same rate as A if L , H_Y , and $\bar{x}$ are fixed. If $\bar{x}$ is fixed, then K must grow at the same rate as A , because total usage of capital is $A\bar{x}\eta$. Let g denote the growth rate of A , Y , and K . Since K/Y is a constant, the ratio

$$\frac{C}{Y} = 1 - \frac{\dot{K}}{Y} = 1 - \frac{\dot{K}}{K} \frac{K}{Y}$$

must also be constant. The common growth rate g for all these variables is therefore

$$g = \frac{\dot{C}}{C} = \frac{\dot{Y}}{Y} = \frac{\dot{K}}{K} = \frac{\dot{A}}{A} = \delta H_A.$$

Together with equation (10), the constraint $H_Y = H - H_A$ implies a relation between the growth rate g and the interest rate r :

$$g = \delta H_A = \delta H - \frac{\alpha}{(1 - \alpha - \beta)(\alpha + \beta)} r, \quad (11)$$

which can be simplified as

$$g = \delta H_A = \delta H - \Lambda r, \quad (11')$$

where Λ is a constant that depends on the technology parameters α and β ,

$$\Lambda = \frac{\alpha}{(1 - \alpha - \beta)(\alpha + \beta)}. \quad (12)$$

Implicitly, the allocation of H between the two sectors is constrained by the requirement that H_A must be nonnegative and that H_Y can be no larger than H . This implies that g is nonnegative. If this constraint is binding, equation (10) will hold as an inequality.

To close the model, it remains to impose the relation between the growth rate g and the interest rate r implied by the preference side of the model, $g = \dot{C}/C = (r - \rho)/\sigma$. Combined with equation (11), this gives an expression for g in terms of the fundamentals of the model,

$$g = \frac{\delta H - \Lambda \rho}{\sigma \Lambda + 1}, \quad (13)$$

where Λ is as defined in equation (12).

The expression for the growth rate suggests a minor technical re-

striction. For the integral in the consumer's preferences to be finite, the rate of growth of current utility $(1 - \sigma)g$ must be less than the discount rate ρ . Thus, for $\sigma \in [0, 1)$, $(1 - \sigma)\delta H/(\Lambda + 1)$ must be less than the discount rate ρ . If this does not hold, the integral can be infinite and some kind of overtaking criterion must be used to describe the behavior of the consumer.

VI. Interpretation and Welfare Properties of the Solution

Almost all the content of the model is contained in equation (11), which summarizes the effects of the technological side of the model, including the effects of imperfect competition in the market for producer durables. Part of the intuition behind equation (11) is easy to grasp. The opportunity cost of human capital is the wage income that can be earned instantaneously in the manufacturing sector. The return to investing human capital in research is a stream of net revenue that a design generates in the future. If the interest rate is larger, the present discounted value of the stream of net revenue will be lower. Less human capital will be allocated to research, and the rate of growth will be lower.

Other aspects of the intuition behind equation (11) are more subtle. The surprising feature of equation (11) is that neither L nor the parameter η is present. Equation (4) shows that an increase in L increases the demand faced by each monopolistic firm selling a durable good. From the monopoly problem (5), it is clear that a reduction in η reduces the costs of the monopolist and increases output $\bar{x}$. In either case, the stream of net revenue generated by a new design increases. Nonetheless, the amount of human capital devoted to research increases neither when L increases nor when η falls.

In a partial equilibrium analysis, one would always expect any change that increases the return to an activity to increase the allocation of resources to that activity. What a general equilibrium analysis emphasizes is that any intervention that increases the return to one activity can very well increase the return to some other activity that competes with the first activity for resources. An increase in L or a reduction in η (which increases $\bar{x}$) raises the return to human capital employed in manufacturing at the same time that it raises the return to human capital in research. For the functional forms used here, these two effects exactly cancel. The allocation of human capital between research and manufacturing does not change when L and η change.

This kind of exact cancellation is not a robust feature of the model.

If slightly different functional forms were used, the net effect of an increase in L or a decrease in η could be either to increase or to reduce the amount of human capital used in the research sector. For example, in an extension of this model (Romer 1990), I show how an increase in L could reduce research effort and the rate of growth. The correct inference from this model is that the effect of an increase in L or a decrease in η on the rate of growth is ambiguous, something that a priori theorizing cannot resolve.

This ambiguity contrasts with the strong and robust implication that reductions in the interest rate will speed up growth. Equivalently, from equation (12), it follows that any change in the preference parameters that acts to reduce the interest rate (an increase in patience captured by a decrease in the discount rate ρ , or an increase in the intertemporal rate of substitution, as captured by a decrease in σ) will increase research and growth. This implication follows directly from the assumption that the benefits of research come largely in the future and that the costs are incurred immediately.

From a policy point of view, the difference between a reduction in η and a reduction in the equilibrium interest rate is very important. A direct subsidy to investment in physical capital financed by a lump-sum tax is mathematically equivalent to a reduction in η . (An illustration of this easy result is given in Romer [1989].) On the basis of an intuition developed for the one-sector model, economists often identify the marginal product of physical capital with the market interest rate. If they were the same, anything that increases the capital stock and reduces the marginal product of physical capital would have the same growth-enhancing effect as a reduction in the interest rate. The intuition that subsidies to physical capital accumulation will also speed up growth in A was reinforced by the first generation of endogenous growth models. In Arrow's (1962a) model of learning by doing and in my first model (Romer 1986), the rate of growth of A was forced by assumption to be the same as the rate of growth of K . As a result, an intervention such as an investment tax credit that increased the accumulation of K necessarily increased the accumulation of A as well.

The model presented here shows that when the decision to invest in physical capital is uncoupled from the decision to invest in research, the effects of a subsidy to physical capital are quite different from the effects of a reduction in the market interest rate. If the fundamental policy problem is that we have too many lawyers and MBAs and not enough engineers, a subsidy to physical capital accumulation is a weak, and possibly counterproductive, policy response.

In the previous paper with differentiated producer durables (Romer 1987), I found that an increase in scale measured by L would increase the rate of growth. Here, an increase in L has no effect.

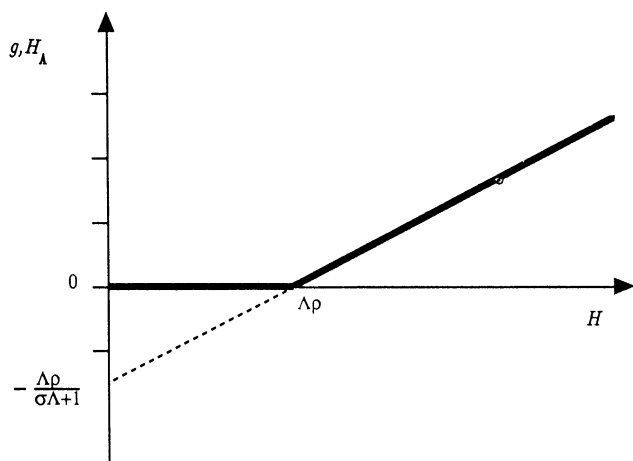


FIG. 2.—Growth rate and amount of human capital in research as a function of total human capital (for $\delta = 1$).

Instead, an increase in scale as measured by total human capital H has the effect of speeding up the rate of growth. This effect is illustrated in figure 2, which plots the rate of growth and the amount of human capital used in research as a function of total human capital. Both of these models have an underlying form of increasing returns in research. As a result, an increase in a scale variable induces an increase in the rate of growth. Human capital is the relevant scale variable in this model because it is the input that is used most intensively in research.

That the research sector in this model exhibits increasing returns is clear from equation (4). If the research technology exhibited constant returns to scale, a doubling of both the human capital and the stock of knowledge would leave the marginal product of human capital in research unchanged. Under the specification used here, a doubling of both leads to an increase in the marginal product of human capital in research. As a result, a permanent increase in the total stock of human capital in the population leads to an increase in the ratio of A to K and a more than proportional increase in the amount of human capital that is devoted to the research sector, as illustrated in figure 2. This implication is of interest from both a historical and a cross-sectional point of view. It is surely the case that the total level of human capital and the fraction of human capital devoted to research are higher now than they were at any time in the past. Moreover, the fraction of human capital devoted to research is apparently highest in the most developed countries of the world.

An extreme possibility illustrated in figure 2 is that if the total level

of human capital is too small, stagnation may arise. If H is too low, the nonnegativity constraint on H_A is binding and growth does not take place. In this case, all the feasible growth rates for A are too small relative to the discount rate to justify the sacrifice in current output necessary for growth to take place. This result offers one possible way to explain the wide variation in growth rates observed among countries and the fact that in some countries growth in income per capita has been close to zero. This explanation is reminiscent of the explanation for the absence of growth in prehistoric time that is offered by some historians and anthropologists: civilization, and hence growth, could not begin until human capital could be spared from the production of goods for immediate consumption. This model cannot offer a complete explanation for these observations because it treats the stock of H (and of L) as given, but it does suggest directions for further work.

The effects of a subsidy to capital can be contrasted with a policy designed to encourage research. A subsidy to employment in the research sector that is financed through lump-sum taxes has the same effects on growth as an increase in the productivity parameter δ in equation (3). In the long run, the subsidy will cause an increase in the growth rate, a fall in P_A , and a reduction in $\bar{x}$ and in the ratio of K to A .

There are two reasons to expect that too little human capital is devoted to research. The most obvious reason is that research has positive external effects. An additional design raises the productivity of all future individuals who do research, but because this benefit is nonexcludable, it is not reflected at all in the market price for designs. The second and an equally important reason why too little human capital is devoted to research is that research produces an input that is purchased by a sector that engages in monopoly pricing. The markup of price over marginal cost forces a wedge between the marginal social product of an input used in this sector and its market compensation. Equation (7) shows that a new design (i.e., an additional unit of A) increases output at every date by an amount $H_Y^\alpha L^\beta \bar{x}^{1-\alpha-\beta}$; from the calculation in (5), the producer of a design captures only a fraction $1 - \alpha - \beta$ of this net benefit to society.

Both of these effects cause human capital to be undercompensated. The marginal product of capital in the manufacturing sector is equal to the wage w_H , but the marginal product in the research sector is higher than the wage because the price of the patent captures only part of the social value of the patent. From equation (9), (10), or (11), it follows that an increase in total human capital would lead to an increase in the amount of human capital employed in the research sector. As a result, in equilibrium, the marginal value of an additional

unit of human capital is higher than the market wage. For simplicity, the model here has treated the stock of total human capital as being exogenously fixed, so undercompensating human capital has no welfare effects. In a more realistic model in which human capital is accumulated endogenously, the supply will be too low. As a result, a second-best policy for a government that cannot affect the allocation of human capital between different sectors would be to subsidize the production of human capital.

Within the confines of the model, the social optimum can be achieved by subsidizing the accumulation of A . Demonstrating this result rigorously starting from arbitrary initial conditions forces the analysis to depart from consideration of balanced growth paths. Any intervention designed to move an economy from one balanced growth path to another must consider the transition dynamics along the way, and an explicit analysis of these dynamics is beyond the scope of this paper. It is easy, however, to compare the growth rate along the balanced growth path that emerges from the equilibrium with the one that would emerge from the solution to a social planning problem. Because of the symmetry between the different producer durables in this economy, the optimal level of $x(i)$ is the same for all i between zero and A . This level is related to K and A by the constraint that $K = \eta x A$. Using this to express x in terms of K and A , we can therefore write the social planning problem for this economy as

$$\max \int_0^\infty \frac{C^{1-\sigma} - 1}{1-\sigma} e^{-\rho t} dt,$$

subject to

$$\dot{K} = \eta^{\alpha+\beta-1} A^{\alpha+\beta} H_Y^\alpha L^\beta K^{1-\alpha-\beta} - C,$$

$$\dot{A} = \delta H_A A,$$

$$H_Y + H_A \leq H.$$

As is shown in the Appendix, the balanced growth solution to the first-order necessary conditions for this problem has a growth rate g^* that is given by

$$g^* = \frac{\delta H - \Theta \rho}{\Theta \sigma + (1 - \Theta)}, \quad (14)$$

where $\Theta = \alpha/(\alpha + \beta)$. The coefficient Λ from equation (12) is equal to this coefficient Θ times the markup from the monopoly sector, $1/(1 - \alpha - \beta)$. This accounts for part of the difference between the equilibrium and the socially optimal rate of growth. However, there is an additional effect because the term in the denominator in equation (14) also replaces the constant 1 from equation (13) with the expres-

sion $1 - \Theta$. This change reflects the effect of correcting for the external effects associated with the production of new ideas. Both of these changes in the expression for the rate of growth—the substitution of Θ for Λ and the substitution of $1 - \Theta$ for 1 —cause the socially optimal allocation of human capital to research to be higher, and this causes the socially optimal rate of growth to be higher.

VII. Growth, Trade, and Research

The final observations about this model pertain to its implications for growth, trade, and research. These can be seen most simply by comparing the balanced growth equilibrium for two identical closed economies that operate in isolation with the balanced growth equilibrium that would obtain if the economies had always been fully integrated. In isolation, the common growth rate is given by g from equation (13), with H set equal to the amount of human capital in each country. In the second case, the growth rate is found by replacing H by total worldwide human capital, $2H$. Both the fraction H_A/H of total worldwide human capital devoted to research and the rate of growth increase.

This thought experiment suggests why a decision to engage in trade may be important even for a country that has a large population, such as China or India. If access to a large number of workers or consumers were all that mattered, having a large population would be a good substitute for trade with other nations. The model here suggests that what is important for growth is integration not into an economy with a large number of people but rather into one with a large amount of human capital. Many of the details of trade between different economies of this kind remain to be worked out,² but since growth seems to be correlated with the degree of integration into worldwide markets but not closely related to population size or density, the results from this model seem promising.

The most direct test of this implication of the model would come from a controlled experiment in which the level of research activity was monitored both before and after a country was opened to trade with the rest of the world. Sokoloff (1988) reports historical data on a natural experiment that comes close to this test. He finds cross-sectional variation that supports the model: counties in the United States in the early nineteenth century that had access to navigable waterways had higher rates of patenting than counties that did not.

² Since the first draft of this paper was written, Grossman and Helpman (1989a, 1989b) have explored detailed models of trade and growth with this kind of underlying technology.

More convincingly, he shows that over time, the introduction of water transportation (because of either the construction of a new canal or the dredging of a river) was followed by a sharp increase in the rate of patenting in counties adjacent to the waterway.

In subsequent work, Sokoloff and Khan (1989) examine the time-series variability in patenting for specific individuals. They find that there was a substantial group of people with broad general knowledge who moved in and out of research in response to aggregate disturbances (primarily business cycles and a trade embargo by the British) just as one would expect if human capital that could be used to do research had alternative uses in manufacturing and commercial trade. They conclude that at least in the nineteenth century, there was a fairly elastic short-run supply of human capital for use in research. Combined with the prior evidence that patenting activity does respond to changes in the size of the market, this offers some assurance that the basic mechanisms described in the model are relevant for historical experience.

VIII. Conclusions

The model presented here is essentially the one-sector neoclassical model with technological change, augmented to give an endogenous explanation of the source of the technological change. The most robust welfare conclusion from the model is that because research projects exchange current costs for a stream of benefits in the future, the rate of technological change is sensitive to the rate of interest. Although all the research is embodied in capital goods, a subsidy to physical capital accumulation may be a very poor substitute for direct subsidies that increase the incentive to undertake research. In the absence of feasible policies that can remove the divergence between the social and private returns to research, a second-best policy would be to subsidize the accumulation of total human capital.

The most interesting positive implication of the model is that an economy with a larger total stock of human capital will experience faster growth. This finding suggests that free international trade can act to speed up growth. It also suggests a way to understand what it is about developed economies in the twentieth century that permitted rates of growth of income per capita that are unprecedented in human history. The model also suggests that low levels of human capital may help explain why growth is not observed in underdeveloped economies that are closed and why a less developed economy with a very large population can still benefit from economic integration with the rest of the world.

Appendix

Calculation of the Balanced Growth Social Optimum

To derive the necessary conditions for the social optimization problem given in the text, construct the current-valued Hamiltonian:

$$\mathcal{H} = \frac{C^{1-\sigma}}{1-\sigma} + \lambda[\eta^{\alpha+\beta-1}A^{\alpha+\beta}(H - H_A)^{\alpha}L^{\beta}K^{1-\alpha-\beta} - C] + \mu\delta H_A A.$$

The necessary conditions follow by maximizing $\mathcal{H}$ with respect to the control variables C and H_A , and from the equations for the evolution of the multipliers λ and μ :

$$\dot{\lambda} = \rho\lambda - \frac{\partial \mathcal{H}}{\partial K}, \quad \dot{\mu} = \rho\mu - \frac{\partial \mathcal{H}}{\partial A}.$$

The first-order condition for maximizing $\mathcal{H}$ with respect to C gives the usual expression relating marginal utility and the multiplier λ :

$$C^{-\sigma} = \lambda. \quad (\text{A1})$$

If the symbol Δ is used to represent the term $\eta^{\alpha+\beta-1}A^{\alpha+\beta}(H - H_A)^{\alpha}L^{\beta}K^{1-\alpha-\beta}$ from the Hamiltonian, the first-order condition for maximizing $\mathcal{H}$ with respect to H_A can be written as

$$\Delta = (H - H_A) \frac{\delta\mu}{\alpha\lambda} A. \quad (\text{A2})$$

Then with equation (A2), the evolution equation for μ can be simplified to yield

$$\frac{\dot{\mu}}{\mu} = \rho - \delta \left(\frac{\alpha + \beta}{\alpha} H - \frac{\beta}{\alpha} H_A \right). \quad (\text{A3})$$

For a balanced growth equilibrium, it must be the case that $\dot{\mu}/\mu = \dot{\lambda}/\lambda$ and that $\dot{C}/C = \dot{A}/A$. With equation (A1), these can be combined to yield $-\sigma(\dot{A}/A) = \dot{\mu}/\mu$. Then combining the evolution equation for A , $\dot{A}/A = \delta H_A$, with equation (A3) gives an equation in H , H_A , and the basic parameters of the model:

$$-\sigma\delta H_A = \rho - \delta \left(\frac{\alpha + \beta}{\alpha} H - \frac{\beta}{\alpha} H_A \right). \quad (\text{A4})$$

Equation (14) in the text can be derived from this equation by solving for H_A and using the fact that the growth rate g is given by $g = \delta H_A$.

References

- Abramovitz, Moses. "Resource and Output Trends in the United States since 1870." *A.E.R. Papers and Proc.* 46 (May 1956): 5–23.
- Arrow, Kenneth J. "The Economic Implications of Learning by Doing." *Rev. Econ. Studies* 29 (June 1962): 155–73. (a)
- . "Economic Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, edited by Richard Nelson. Princeton, N.J.: Princeton Univ. Press (for NBER), 1962. (b)

- Becker, Gary S.; Murphy, Kevin M.; and Tamura, Robert. "Human Capital, Fertility, and Economic Growth." *J.P.E.*, this issue.
- Cornes, Richard, and Sandler, Todd. *The Theory of Externalities, Public Goods, and Club Goods*. Cambridge: Cambridge Univ. Press, 1986.
- Dasgupta, Partha, and Stiglitz, Joseph E. "Learning-by-Doing, Market Structure and Industrial and Trade Policies." *Oxford Econ. Papers* 40 (June 1988): 246–68.
- Dixit, Avinash K., and Stiglitz, Joseph E. "Monopolistic Competition and Optimum Product Diversity." *A.E.R.* 67 (June 1977): 297–308.
- Ethier, Wilfred J. "National and International Returns to Scale in the Modern Theory of International Trade." *A.E.R.* 72 (June 1982): 389–405.
- Gollop, Frank M., and Jorgenson, Dale W. "U.S. Productivity Growth by Industry." In *New Developments in Productivity Measurement and Analysis*, edited by John W. Kendrick and Beatrice N. Vaccara. Chicago: Univ. Chicago Press (for NBER), 1980.
- Griliches, Zvi. "Issues in Assessing the Contribution of Research and Development to Productivity Growth." *Bell J. Econ.* 10 (Spring 1979): 92–116.
- Grossman, Gene M., and Helpman, Elhanan. "Comparative Advantage and Long Run Growth." Working Paper no. 2809. Cambridge, Mass.: NBER, January 1989. (a).
- . "Growth and Welfare in a Small Open Economy." Working Paper no. 2970. Cambridge, Mass.: NBER, May 1989. (b)
- . "Product Development and International Trade." *J.P.E.* 97 (December 1989): 1261–83. (c)
- Heckman, James J. "A Life-Cycle Model of Earnings, Learning, and Consumption." *J.P.E.* 84, no. 4, pt. 2 (August 1976): S11–S44.
- Helpman, Elhanan, and Krugman, Paul R. *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*. Cambridge, Mass.: MIT Press, 1985.
- Jones, Larry, and Manuelli, Rodolfo. "A Model of Optimal Economic Growth." Working paper. Stanford, Calif.: Stanford Univ., 1988.
- Jorgenson, Dale W.; Gollop, Frank M.; and Fraumeni, Barbara M. *Productivity and U.S. Economic Growth*. Cambridge, Mass.: Harvard Univ. Press, 1987.
- Judd, Kenneth L. "On the Performance of Patents." *Econometrica* 53 (May 1985): 567–85.
- Kendrick, John W. "Productivity Trends: Capital and Labor." *Rev. Econ. and Statis.* 38 (August 1956): 248–57.
- King, Robert G., and Rebelo, Sergio. "Business Cycles with Endogenous Growth." Working paper. Rochester, N.Y.: Univ. Rochester, 1987.
- Lucas, Robert E., Jr. "On the Mechanics of Economic Development." *J. Monetary Econ.* 22 (July 1988): 3–42.
- Maddison, Angus. *Phases of Capitalist Development*. Oxford: Oxford Univ. Press, 1982.
- Nordhaus, William D. *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change*. Cambridge, Mass.: MIT Press, 1969.
- Pascoa, Mario. "Monopolistic Competition and No-Neighboring-Goods." Working Paper no. 86-14. Philadelphia: Univ. Pennsylvania, Center Analytic Res. Econ. and Soc. Sci., 1986.
- . "Noncooperative Equilibrium and Chamberlinian Monopolistic Competition." *J. Econ. Theory* (1990), in press.
- Rebelo, Sergio. "Long Run Policy Analysis and Long Run Growth." Working paper. Rochester, N.Y.: Univ. Rochester, 1988.

- Romer, Paul M. "Increasing Returns and Long-Run Growth." *J.P.E.* 94 (October 1986): 1002–37.
- . "Growth Based on Increasing Returns Due to Specialization." *A.E.R. Papers and Proc.* 77 (May 1987): 56–62.
- . "Capital Accumulation in the Theory of Long Run Growth." In *Modern Business Cycle Theory*, edited by Robert J. Barro. Cambridge, Mass.: Harvard Univ. Press, 1989.
- . "Capital, Labor, and Productivity." *Brookings Papers Econ. Activity* (1990), in press.
- Rosen, Sherwin. "A Theory of Life Earnings." *J.P.E.* 84, no. 4, pt. 2 (August 1976): S45–S67.
- Schumpeter, Joseph A. *Capitalism, Socialism, and Democracy*. New York: Harper, 1942.
- Shell, Karl. "Toward a Theory of Inventive Activity and Capital Accumulation." *A.E.R. Papers and Proc.* 56 (May 1966): 62–68.
- . "A Model of Inventive Activity and Capital Accumulation." In *Essays on the Theory of Optimal Economic Growth*, edited by Karl Shell. Cambridge, Mass.: MIT Press, 1967.
- . "Inventive Activity, Industrial Organisation and Economic Growth." In *Models of Economic Growth*, edited by James A. Mirrlees and Nicholas H. Stern. New York: Wiley, 1973.
- Sokoloff, Kenneth L. "Inventive Activity in Early Industrial America: Evidence from Patent Records, 1790–1846." *J. Econ. Hist.* 48 (December 1988): 813–50.
- Sokoloff, Kenneth L., and Khan, Zorina. "The Democratization of Invention during Early Industrialization: Evidence from the United States, 1790–1846." Working paper. Los Angeles: Univ. California, 1989.
- Solow, Robert M. "A Contribution to the Theory of Economic Growth." *Q.J.E.* 70 (February 1956): 65–94.
- . "Technical Change and the Aggregate Production Function." *Rev. Econ. and Statis.* 39 (August 1957): 312–20.
- Uzawa, Hirofumi. "Optimum Technical Change in an Aggregative Model of Economic Growth." *Internat. Econ. Rev.* 6 (January 1965): 18–31.
- Wilson, Robert. "Informational Economies of Scale." *Bell J. Econ.* 6 (Spring 1975): 184–95.