The O-Ring Theory of Economic Development
Author(s): Michael Kremer
Source: *The Quarterly Journal of Economics*, Vol. 108, No. 3 (Aug., 1993), pp. 551-575
Published by: The MIT Press
Stable URL: http://www.jstor.org/stable/2118400
Accessed: 26/10/2010 18:09

THE O-RING THEORY OF ECONOMIC DEVELOPMENT*

MICHAEL KREMER

This paper proposes a production function describing processes subject to mistakes in any of several tasks. It shows that high-skill workers—those who make few mistakes—will be matched together in equilibrium, and that wages and output will rise steeply in skill. The model is consistent with large income differences between countries, the predominance of small firms in poor countries, and the positive correlation between the wages of workers in different occupations within enterprises. Imperfect observability of skill leads to imperfect matching and thus to spillovers, strategic complementarity, and multiple equilibria in education.

Many production processes consist of a series of tasks, mistakes in any of which can dramatically reduce the product's value. The space shuttle *Challenger* had thousands of components: it exploded because it was launched at a temperature that caused one of those components, the O-rings, to malfunction. "Irregular" garments with slight imperfections sell at half price. Companies can fail due to bad marketing, even if the product design, manufacturing, and accounting are excellent. This paper argues that the analysis of such processes can help explain several stylized facts in development and labor economics.

The first section of the paper proposes a production function in which production consists of many tasks, all of which must be successfully completed for the product to have full value. I assume that it is not possible to substitute several low-skill workers for one high-skill worker, where skill refers to the probability a worker will successfully complete a task. Subsection I.1 solves for equilibrium wages as a function of worker skill under this production function, and shows that firms will match together workers of similar skill. Subsection I.2 argues that this production function is consistent with a series of stylized facts in development and labor economics, including the enormity of wage and productivity differences between rich and poor countries and the positive correlation between wages of workers in different occupations within firms. A variant of the model in which tasks are performed sequentially implies that the share of agriculture in GNP will fall with development.

*I am grateful to Roland Benabou, Kala Krishna, Eric Maskin, Paul Romer, Sherwin Rosen, Xavier Sala-i-Martin, Philippe Weil, many former classmates at Harvard, and especially Robert Barro for comments and suggestions. The National Science Foundation provided financial support during the period this paper was written.

The Quarterly Journal of Economics, August 1993

Subsection I.3 shows that higher skill workers will use more complex technologies that incorporate more tasks. This may help explain why household production and small firms are the dominant form of industrial organization in developing countries and why there is a positive correlation between wages and firm size within countries. Although each of these stylized facts may be due to a variety of factors, taken together, they suggest that this type of production function is empirically relevant.

Section II endogenizes worker skill as the product of investment in human capital. If workers are matched together perfectly, investment in human capital will be Pareto optimal. However, if worker skill cannot be perfectly observed, so matching is imperfect, there will be underinvestment in human capital, so education subsidies will be optimal. Moreover, there will be strategic complementarity in this investment, so these subsidies will have multiplier effects, and small differences between countries in exogenous factors will cause large differences in worker skill. If strategic complementarity is strong enough, there will be multiple equilibria.

Section III generalizes the argument to production functions with arbitrary returns to scale, and discusses extensions to production functions which induce firms to match together workers of dissimilar skill and to production functions in which different tasks enter the production function asymmetrically. A conclusion summarizes the results.

The model builds on Rosen's [1981] analysis of superstars and on Rosen [1982], Miller [1983], and Lucas [1978], which build models of organizational hierarchy in which managerial skill enters the production function multiplicatively, and unskilled labor enters with standard diminishing returns. In these models, agents with skill below some cutoff level become workers, and agents with skill above the cutoff level become managers. Higher skill managers supervise more employees. This paper differs in examining skill interaction among workers at the same level of hierarchy. In this model, rather than supervising more employees, high-skill agents are matched with high-skill coworkers.[1] This paper thus combines Rosen's analysis of multiplicative quality effects with Becker's [1981] analysis of matching in marriage markets. It is also related to the work of Sah and Stiglitz [1985, 1986] and Sobel [1992] in applying the literature on reliability to organizations.

1. An appendix available from the author works out an example in which higher quality workers have both more subordinates and higher skill coworkers.

I. The O-Ring Production Function and Applications

I.1. The O-Ring Production Function

Consider a firm using a production process consisting of n tasks. For example, in an automobile factory one task might be installing the brakes, and in a restaurant one task might be waiting on tables. For simplicity of exposition, I assume that each task requires a single worker, but this need not be true in general, and n should be taken as referring to the number of tasks, not the number of workers. For now I shall assume that n is technologically fixed. Firms can replicate the production process an arbitrary number of times. A worker's skill (or quality) at a task, q, is defined by the expected percentage of maximum value the product retains if the worker performs the task. Thus, a q of 0.95 could refer to a worker who has a 95 percent chance of performing the task perfectly and a 5 percent chance of performing it so badly the product is worthless, to a worker who always performs a task in such a way that the product retains 95 percent of its value, or to a worker who has an 50 percent chance of performing the task perfectly and a 50 percent chance of making a mistake that reduces the value of the product to 90 percent of its maximum possible value. The probability of mistakes by different workers is independent. Capital k enters the production function in conventional Cobb-Douglas form and is not differentiated by quality. Define B as output per worker with a single unit of capital if all tasks are performed perfectly. Expected production is thus

$$E(y) = k^{\alpha}(\Pi_{i=1}^{n} q_i)\, nB. \tag{1}$$

Firms are risk-neutral, so the remainder of the paper drops the distinction between production and expected production. There is a fixed supply of capital, k^*, and a continuum of workers following some exogenous distribution of quality, $\phi(q)$. Workers face no labor-leisure choice and supply labor inelastically.

This O-ring production function differs from the standard efficiency units formulation of labor skill, in that it does not allow quantity to be substituted for quality within a single production chain. For example, it assumes that it is impossible to substitute two mediocre advertising copywriters, chefs, or quarterbacks for one good one.[2] The particular functional form set forth in this section exhibits increasing returns to the skill of the workforce

2. This production function is similar to that in Stinchcombe and Harris [1969].

taken as a whole, but as Section III discusses, much of the analysis generalizes to symmetric production functions with a positive cross derivative in worker skill.

It is possible to solve for a competitive equilibrium—defined as an assignment of workers to firms, a set of wage rates, $w(q)$, and a rental rate r, such that firms maximize profits and the market clears for capital and for workers of all skill levels.

Firms facing a wage schedule, $w(q)$, and a rental rate, r, choose a level of capital, k, and the skill of each worker, q_i, to maximize revenue minus cost:

$$\max_{k,\{q_i\}} k^{\alpha} (\Pi_{i=1}^{n} q_i)\, nB - \sum_{i=1}^{n} w(q_i) - rk. \tag{2}$$

The first-order condition associated with each of the q_i is

$$\frac{dw(q_i)}{dq_i} = \frac{dy}{dq_i} = (\Pi_{j \neq i}\, q_j)\, nBk^{\alpha}. \tag{3}$$

Thus, the increase in output a firm obtains by replacing one worker with a slightly higher skill worker while leaving the skill of its other workers unchanged must equal the increase in its wage bill necessary to pay the higher skill worker. The marginal product of skill, dy/dq_i, must equal the marginal cost of skill, $dw(q_i)/dq_i$, or else the firm would prefer to employ either lower or higher skill workers.

The search for equilibria can be restricted to those allocations of workers to firms in which all workers employed by any single firm have the same q. This is because the derivative of the marginal product of skill for the ith worker with respect to the skill of the other workers is positive:

$$\frac{d^2y}{dq_i d(\Pi_{j \neq i}\, q_j)} = nBk^{\alpha} > 0. \tag{4}$$

This positive cross derivative means that firms with high q workers in the first $n - 1$ tasks place the highest value on having high-skill workers in the nth task, so they bid the most for these workers. Thus, in equilibrium, workers of the same skill are matched together in firms, just as marriage partners of similar quality are matched together in Becker's [1981] marriage model.[3]

3. Becker [1981, p. 72] reproduces a formal proof by William Brock which shows that a positive cross derivative implies positive assortative matching. See also Sattinger [1975].

For now I assume perfect matching; Section II examines imperfect matching.

Given that workers of the same skill are matched together, $q_i = q_i$ for all j and the first-order condition on q can be rewritten as

$$\frac{dw}{dq} = q^{n-1}nBk^{\alpha}. \tag{5}$$

The first order condition on capital, $\alpha k^{\alpha-1}q^{n}nB = r$, implies that

$$k = \left(\frac{\alpha q^{n}nB}{r}\right)^{1/(1-\alpha)}. \tag{6}$$

It is straightforward to show that payments to capital are αy. The equilibrium rental rate on k, r, will be that which equates the supply of capital, k^*, with the demand, which is given by summing up the capital demanded by the firms hiring all the different skill levels of workers, from zero to one. Since the density of firms hiring workers of a particular skill is $1/n$ times the density of workers of that skill level, this implies that

$$\int_0^1 \left(\frac{\alpha q^{n}nB}{r}\right)^{1/(1-\alpha)} \frac{1}{n}\,\partial\phi(q) = k^*. \tag{7}$$

Thus, $r = \alpha Bn^{\alpha}[\int_0^1 q^{n/(1-\alpha)}\partial\phi(q)/R^*]^{1-\alpha}$.
(Alternatively, in an open economy, r would be fixed, and k^* would be the equilibrium level of capital.) The first-order condition on q, (5), can be rewritten by substituting in the value of k from equation (6):

$$\frac{dw}{dq} = nq^{n-1}B\left(\frac{\alpha q^{n}nB}{r}\right)^{\alpha/(1-\alpha)}. \tag{8}$$

Integrating generates the set of wage schedules that allows firms hiring workers of any single level of skill to satisfy this first-order condition:

$$w(q) = (1-\alpha)(q^{n}B)^{1/(1-\alpha)}\left(\frac{\alpha n}{r}\right)^{\alpha/(1-\alpha)} + c, \tag{9}$$

or equivalently,

$$w(q) = (1-\alpha)q^{n}Bk^{\alpha} + c. \tag{10}$$

The constant of integration, c, represents the wage of a worker of skill zero, who never performs a task successfully. Multiplying the wage schedule by n, the number of workers, shows that the total wage bill is $(1 - \alpha)Y + nc$. Since payments to capital are αY, the zero profit condition implies that the constant of integration must equal zero.

Since profits are zero for all firms given the wage schedule $w(q)$, firms are indifferent as to the skill level of their workers as long as their labor force is of homogenous skill. Equilibrium holds when firms demand the number of workers of each skill available in the population. Since this is a well-behaved problem, this competitive equilibrium is optimal and unique up to reassignments of workers of equal skill.

I.2. Applications to Development and to Labor Markets

O-ring production functions are consistent with a series of stylized facts in development and labor economics. While each of these stylized facts may be due to a variety of different factors, taken together, they suggest that O-ring production functions are empirically relevant.

1. Wage and productivity differentials between rich and poor countries are enormous.

According to the World Bank [1990], United States GDP per capita is twenty times that of Bangladesh using purchasing-power-parity adjusted figures, and more than 100 times that of Bangladesh using exchange rate valuations, which presumably indicate ability to produce tradable goods. Either way, the disparity is enormous. Differences in physical capital have been used to explain international income differences, but as Lucas [1990] argues, physical capital should be mobile given large enough incentives. Lucas calculates that if the income difference between the United States and India were due to differences in physical capital alone, the marginal product of capital in India would be 58 times that of the United States.

Worker quality could be another potential source of differences in income levels. Barro [1991] and Mankiw, Romer, and Weil [1992] find that human capital is an important factor in economic growth. Moreover, microeconomic studies find astonishingly large differences between countries in worker productivity: Clark [1987] examines early twentieth century textile mills and finds that "... one New England cotton textile operative performed as much work as 1.5 British, 2.3 German, and nearly 6 Greek, Japanese, Indian, or Chinese workers." Noting that the same equipment was used worldwide, Clark rules out differences in technology and capital intensity as causes of these productivity differences, and points to differences in "personal efficiency" between workers in different countries. However, even if some national differences in worker skill were plausible, it would be difficult to understand what could cause differences of these magnitudes.

An O-ring production function provides a mechanism through which small differences in worker skill create large differences in productivity and wages. As is clear from equation (9), under this production function, equilibrium wages are homogenous of degree $n/(1 - \alpha)$ in q, so small differences in worker skill create large differences in output and wages. Moreover, in equilibrium more physical capital is used with higher skill workers, thus helping answer Lucas' question about why capital does not flow from rich to poor countries. Intuitively, higher skill workers are less likely to make mistakes that waste the rental value of capital, and it is therefore optimal for them to use more capital.

2. Firms hire workers of different skill and produce different quality products.

In many industries different firms hire different qualities of workers. Restaurants, for example, come in a range of quality levels. McDonald's does not hire famous chefs, and Maxim's does not hire teenage waiters. Charlie Parker and Dizzy Gillespie work together, and so do Donny and Marie Osmond. For tradable goods this division is often international, creating implications for both development and labor markets. Italy, Taiwan, and China all export bicycles. Perhaps part of what allows Italian companies to compete with cheaper Chinese labor is substitution of cheaper Italian capital. But an argument similar to Lucas' indicates that tremendous differences in the cost of capital would be needed to equalize production costs between Italy and China. Systematic differences in product quality, associated with differences in the skill of the employees, are a more plausible explanation of why Italian bicycle manufacturers can compete with their Chinese counterparts.

3. There is a positive correlation among the wages of workers in different occupations within enterprises.

Secretaries working for investment banks or major law firms earn more than secretaries working in retail banks or local law offices.[4] Pressures for intrafirm equity and industry rents have been suggested as explanations; but O-ring production functions provide another explanation, since they imply that the highest q secretaries will work with the highest q lawyers and bankers.

4. Firms only offer jobs to some workers rather than paying all workers their estimated marginal product.

Under a conventional production function, a construction firm, for example, could hire bricklayers of any skill and pay according to estimated future output. Under an O-ring production function, the firm needs bricklayers whose skill matches that of its carpenters, electricians, and plumbers. The firm will therefore be willing to expend resources interviewing a number of employees for a single position to find a bricklayer of the right skill. Although the firm could offer to hire a bricklayer of inappropriate skill, this would be pointless since the wage would be far from what the worker could earn elsewhere, and might even be negative. O-ring production functions thus help provide a rationale for job search theories of unemployment, such as Jovanovic [1979], in which workers have different productivity at different firms.

5. Income distribution is skewed to the right.

The model fits the distribution of income, at least to the extent that one believes fundamental parameters are distributed symmetrically. Under the model, if q is distributed symmetrically, y will be skewed to the right, and log y will be symmetric.[5] In fact, the distribution of income is skewed to the right, both within and between countries. The log of income is distributed approximately symmetrically.

I.3. Sequential Production

So far, I have assumed that all tasks are performed simultaneously. In fact, some production processes consist of several stages, undertaken with a technology that allows workers to detect mistakes and avoid wasting further work on defective items. For example, one of Rembrandt's assistants would prepare the canvas, another would paint in most of a figure, and finally, if that were acceptable, Rembrandt would paint the face and hands. As Sobel [1992] has demonstrated in a similar framework, in such processes

4. See Katz and Summers [1989, Table III] for evidence that janitors and secretaries earn more in industries where the average wage is higher.

5. I am grateful to Sherwin Rosen for pointing this out to me.

the highest q workers are allocated to the later stages of production in equilibrium since mistakes there destroy higher valued inputs than in earlier stages. This is similar to hierarchy models, in that one higher stage worker works with more than one lower stage worker.

To see this more formally, assume that several stages of production are needed to transform some free, unproduced primary good into a final good. Each stage requires one unit of labor and one unit of output from the previous stage of production. All stages of production could be conducted within a single firm, or each stage could be conducted in a separate firm. To simplify, I shall henceforth assume that $\alpha = 0$, so capital does not enter the production function. A worker of skill q successfully transforms a unit of the i − 1st stage good into the ith stage good with probability q and makes a mistake that destroys the product with probability $1 - q$.[6] Let p_i denote the price of the good at the ith stage of production. Expected profits for a firm in the ith stage of production employing one worker of skill q_i and using one unit of the i − 1st good as an input are $q_i p_i - p_{i-1} - w(q_i)$. In equilibrium firms earn zero profits, and therefore $q_i p_i - p_{i-1} - w(q_i) = 0$. This implies that $w(q_i) = q_i p_i - p_{i-1}$, and since $p_i > p_j$ for $i > j$, the equilibrium wage schedule is steeper in q at later stages of production.

Suppose that there were an equilibrium allocation of workers to tasks in which $q_i < q_j$ for $i > j$, that is, in which a higher stage of production had a lower skill worker. Since $i > j, p_i > p_j$. Given the wage schedule derived earlier, if the two workers switched jobs, their total income would change by $(p_i - p_j)(q_j - q_i)$. Since both these terms are positive, total income increases if the workers switch jobs, and hence the allocation of workers to tasks is not an equilibrium. Hence in equilibrium, higher q workers must be allocated to later stages of production.

This variant of the production function is consistent with the following two stylized facts.

6. Poor countries have higher shares of primary production in GNP.
7. Workers are paid more in industries with high value inputs.

Under sequential production, countries with high-skill work-

6. Unlike the rest of the paper, this variant of the production function assumes that workers perform their task either perfectly or so badly that the product is worthless and cannot be used in further stages of production.

ers specialize in products that require expensive intermediate goods, and countries with low-skill workers specialize in primary production. In fact, poor countries have a consistently high share of agriculture and primary production in GNP, even when their land endowments are small. El Salvador, for example, has only one-twelfth of Canada's endowment of arable land per capita, yet its share of agriculture in GDP is 19 percent, compared with 3 percent in Canada. Since Salvadorans are poorer, it is not surprising that they have a larger share of food in consumption, but given the possibility of trade, it is not clear why they have a larger share of agriculture in production. Although other explanations, such as low human capital intensity in agriculture, have been posited, the sequential production model may provide part of the explanation of why poor countries concentrate on primary production. Kwon [1992] finds that productivity in the former Soviet Union lagged the most relative to the United States in final and intermediate goods industries and was closest to United States levels in primary goods industries. (Agriculture was an exception, but this may have been due to worker monitoring problems that make agriculture highly unsuited to state ownership.) Within countries, sequential production helps explain why automobile workers, diamond cutters, and others who work with high value inputs are highly paid.

In addition to fitting the stylized facts above, O-ring production functions increase the quantitative importance of efficiency wages, bottlenecks, and trade restrictions. O-ring production functions strengthen efficiency wage effects because they magnify the loss from shirking. They increase the impact of bottlenecks not only directly, but also indirectly, through their impact on incentives to invest in skill. To see this, assume, for example, that n tasks are required to produce a good, and, taking q as task-specific, consider the effect of halving the q of all the economy's workers in two tasks, say machine maintenance and accounting. Assignments of workers to firms do not change, because the highest q people in the last two tasks are still matched with the highest q people in the first $n - 2$ tasks. Production, however, falls by 75 percent. Moreover, the marginal product of quality, dw/dq_i, falls by 75 percent in the other $n - 2$ sectors and hence so does the incentive to invest in q (through education, for example). As workers in these sectors reduce their investment in skill, they further reduce the level of q in the economy, and thus the incentive to accumulate skill.

Although bottlenecks generate high returns to the missing skills, the market may not remove bottlenecks caused by low quality inputs of public goods such as police protection, electricity and water, or communication and transportation infrastructure. More generally, low domestic capacity in sectors where trade is costly or impossible can create bottlenecks. As Clague [1991a, 1991b] points out, enterprises may become vertically integrated to avoid using unreliable inputs from other parts of the economy. Thus, Chinese factories provide schools and housing for their workers, and western multinationals working in developing countries import some requirements and set up enclave economies to provide others. For example, in Russia, McDonald's could not buy the quality of beef it needed domestically, was not allowed to import it, and therefore arranged its own beef production. However, in becoming vertically integrated, firms are prone to a breakdown anywhere along a longer production chain.

The current view in development economics is that trade restrictions cause large welfare losses, rather than the proverbial small Harberger triangles. O-Ring production functions provide support for this view, since they indicate that trade restrictions, especially quantitative restrictions, can paralyze production by preventing bottleneck sectors from being bypassed.

I.4. Equilibrium Choice of Technology

So far, I have taken n, the number of tasks, as technologically fixed, but the analysis can be generalized to allow firms to choose among technologies with different n. A VCR manufacturer could build anything from a simple \$150 VCR player to an \$800 machine with timer, remote control, and automatic commercial cutting. A farmer could scatter seeds and wait for them to grow or could build terraces, dig irrigation ditches, grow seedlings in a nursery, apply fertilizer to his fields, and hedge risks on the futures market. More fundamentally, firms can choose whether to produce complex products such as aircraft, or simpler products, such as textiles. To simplify, I assume that all tasks require the same amount of labor and define $B(n)$ as the value of output per task if all tasks are performed perfectly. I assume that if all tasks are performed correctly there are benefits to using more complex techology, at least over some range, but that these benefits diminish as technology becomes more complex, so that $B'(0) > 0$ and $B''(n) < 0$.[7]

By increasing n, I do not mean subdividing existing tasks

7. It is possible to replicate the argument below with n restricted to integers, using integer analogues of the assumptions that $B'(0) > 0$ and $B''(n) < 0$.

through Smithian division of labor: there is no reason to assume that the chance of a mistake increases if one worker puts on the bolt and another puts on the nut. Rather, increasing n means switching to different production techniques or products in which there are more potential areas for mistakes that affect the value of the product as a whole. Thus, making a car involves more tasks than making a bicycle, because there are more things that can go wrong. It is more difficult to make cardinal statements about the number of tasks. For example, a waiter can be thought of as performing a single task with a q of 0.97, or three tasks—taking the order, serving the food, and collecting the check—each with a q of approximately 0.99.

In choosing the technology, firms face the problem:

$$\max_{n,\{q_i\}} \left(\Pi_{i=1}^{n} q_i\right) nB(n) - \sum_{i=1}^{n} w(q_i). \tag{11}$$

In equilibrium each firm must satisfy a first-order condition for optimal choice of n and each of the q_i. Since the first-order condition on choice of the q_i is the same as in subsection I.1, the search for equilibria can again be restricted to allocations of workers to firms in which workers of the same skill are matched together, and the firm's problem can be written as

$$\max_{n,q} q^n nB(n) - nw(q). \tag{12}$$

The first-order condition on choice of n is therefore

$$q^n B(n) - w(q) + n[\log(q)q^n B(n) + B'(n)q^n] = 0. \tag{13}$$

The first-order condition on q implies that $w(q) = q^n B(n)$, as in subsection I.1. Substituting for $w(q)$ and simplifying,

$$-\log(q) = B'(n)/B(n). \tag{14}$$

The left-hand side declines monotonically in q. Since $B'(0) > 0$ and $B''(n) < 0$, the right-hand side declines monotonically in n as long as $B'(n) > 0$. Therefore, n is an implicit function of q with $n'(q) > 0$. Hence, firms producing products or using technologies requiring high n will employ high q workers. Intuitively, mistakes are more costly to firms with high n, so they place higher value on skilled workers, and are allocated these workers in equilibrium.

In a more general model in which the products of technologies with different n were imperfect substitutes, the assignment of a worker to a technology would depend not only on his own q, but also on the distribution of q in the economy. For example, if the highest q workers were assigned to the aircraft industry and a particular country had a large supply of high q workers, it might have a higher cutoff level of q above which people worked in the aircraft industry.

The relationship between n and q fits the next stylized fact.

8. Rich countries specialize in complicated products.

The prediction that countries with high q will use technologies requiring more tasks fits the pattern of international specialization in which rich countries specialize in complicated products, such as aircraft, and poor countries produce simpler products such as textiles. One measure of product complexity is the number of different inputs, and Clague [1991a, 1991b] finds that poor countries are relatively less efficient in industries with a large number of input sectors and a high dispersion of input shares, as measured by the U. S. input-output table.

Strictly speaking, n refers to the number of tasks rather than the number of workers. In the absence of a fully worked out theory of the firm, it is difficult to make strong statements about the relationship between a firm and a production process. Nonetheless, if workers improve their efficiency by specializing in particular tasks and if there are a span of control problems in replicating a production process indefinitely and transaction cost problems in dividing it up arbitrarily, then there is likely to be a positive correlation between the number of tasks and the number of workers. Given such a positive correlation, the model is consistent with the following stylized facts.

9. Firms are larger in rich countries.
10. Firm size and wages are positively correlated.

The model predicts that firms in poorer countries will choose lower n technologies, and if there is a correlation between n and firm size, this implies that firms will be smaller in poorer countries. In fact, firms consisting of a single household predominate in most poor countries. This reflects not only the higher share of agriculture in developing countries, but also the structure of firms within sectors. In food retailing, for example, firms in developing countries typically consist of a single person or household, whereas rich countries have giant supermarket chains with specialized cashiers, stockers, truckers, and advertising copywriters. Clague [1991a, 1991b] finds that rich countries have higher relative efficiency in industries with more employees per firm. Within countries, the model's implication that higher n firms will employ higher q workers matches the empirical correlations between firm size and observable indicators of worker quality, and between firm size and wages, documented by Brown and Medoff [1989], among others.

II. Endogenizing Skill Under Imperfect Information

Section I took q as exogenous and argued that small differences in q can have important effects. This section endogenizes skill as the product of investment in education or effort, e, in order to model possible sources of skill differences. If workers can match perfectly—that is, if workers can be matched with others of similar skill no matter what their choice of skill—they will face the wage schedule derived in Section I, and will therefore choose skill optimally.[8] As mentioned above, workers in different countries might choose different q due to differences in education systems, tax policies, or nontradable bottleneck sectors that affect incentives to invest in education. Under O-ring production functions output is a convex function of q, so the accumulation problems faced by workers may be nonconvex even if q is a concave function of education. As others have shown, if capital markets are imperfect, nonconvex accumulation problems can lead to multiple equilibria in levels of human capital [Dechert and Nishimura, 1983], and to Kuznets curves [Galor and Zeira, 1989]. While interesting dynamics can thus arise even under perfect matching, this paper focuses on incentives for accumulation of human capital when workers are

8. With a continuum of workers and perfectly observable skill, the wage schedule derived in Section I is the unique competitive equilibrium, and it will induce optimal investment in education. However, there could, hypothetically, be additional, suboptimal, Nash equilibria and strategic complementarity in selection among these equilibria. The additional equilibria arise because if all workers chose the same level of skill there would be missing markets for other skill levels. Workers considering choosing skill levels other than the one chosen by all other agents would therefore face lower than competitive equilibrium wages, and hence the economy could coordinate on a suboptimal level of skill (zero, for example). However, I believe that it is unrealistic to focus on these additional equilibria. Doing so is analogous to claiming that there could be an equilibrium in which neither of two complementary goods is produced due to the absence of the other good. These equilibria are fragile because a small number of people could form a self-enforcing agreement to choose the optimal level of education. Moreover, if there were a small error term creating heterogeneity in skill, workers over a range of skill levels would find identical partners and therefore receive the competitive equilibrium wage schedule. Assuming that the net payoff to education was concave under the competitive equilibrium wage schedule, workers would always receive a higher payoff by choosing a level of education slightly closer to the competitive equilibrium than that chosen by other agents, and this would eliminate any suboptimal equilibria.

imperfectly matched, and hence do not face the wage schedule of Section I, and need not choose the socially optimal e. It first shows that imperfect matching due to limited availability of workers of certain skill levels can lead to the formulation of specialized cities that will be especially attractive to people with high human capital. It then shows that imperfect matching due to imperfect observability of skill leads to underinvestment in skill, strategic complementarity in that investment, and the possibility of multiple equilibria.

In a finite population, in which skill is determined by education and a random error term with a continuous distribution, workers will not be able to match perfectly. Instead they will match in rank order of skill, with the division of a firm's output among its heterogeneous workers determined by a complex bargaining problem. Since skill is often industry- and task-specific, large populations may be needed for people to find close matches in their field. Fred Astaire was born in Omaha, Nebraska; Ginger Rogers, in Independence, Missouri. They had to go to New York to meet each other. Matching thus creates incentives for people to cluster in cities. If there are congestion costs, it may be efficient for tradable sectors to concentrate in different cities: autos in Detroit, fashion in Milan, country music in Nashville.

Under imperfect matching, the marginal product of skill, $dw/dq_i = E(\Pi_{j \neq i} q_j)$, increases with population. With a larger population the coworkers' skills are likely to be closer together, and thus the expectation of this product will be greater.[9] Thus, technological advances that allow matching between different regions or political or cultural changes that allow matching between different groups will increase not only production, but also incentives to invest in human capital. The greater return to q in areas with high population may help explain why educated people are more likely to migrate from rural areas to cities.

One problem with using differences in worker skill as an explanation of international income differences is that income differences between countries are large relative to those within countries, and it is unclear why skill differences between countries would be large relative to differences within countries. Imperfect matching provides a partial explanation, since it reduces variation in income within countries relative to variation in average income between countries. To see this, assume that q is a function of

9. This assumes that each worker has more than one coworker, or equivalently, that firms consist of three or more workers.

education and that choice of education depends both on individual-specific factors, such as tastes for education, and country-specific factors, such as taxes on labor income. Further, assume that the distribution of q within each country due to individual-specific factors is thicker in the center than in the tails. Agents then have an incentive to choose a level of investment in education that puts them nearer the center of the distribution, where closer matches are available. This process is self-reinforcing, because as agents choose levels of education that put them near the center of the distribution the tails become even thinner.

Imperfect information about skill is another, probably more important, cause of imperfect matching. Below, I argue that imperfect information leads to underinvestment in skill and to strategic complementarity in this investment. Intuitively, it is more valuable to be a high-skill worker if one has high-skill coworkers, and under imperfect matching, the expected skill of one's coworkers increases in the level of education chosen by the rest of the population. Hence each worker has more incentive to choose a high level of education if other workers choose a high level of education. This creates multiplier effects: for example, a small education subsidy can create large differences in q and production. Sufficiently strong strategic complementarity creates multiple equilibria in investment in skill. To see these effects more formally, suppose that the production technology is

$$Y = n \, \Pi_{i=1}^{n} q_i, \tag{15}$$

where n is fixed, and there is a stochastic education technology such that skill q depends on e, education or effort,

$$\log(q) = \log[g(e)] + \epsilon \qquad \epsilon \sim N(0,\sigma_\epsilon^2), \tag{16}$$

where

$$g' > 0 \qquad g'' < 0 \qquad g'(0) = \infty \qquad g(e) > 0 \qquad g(\infty) = 1. \tag{17}$$

Skill is observed (even by the worker himself) only through a test score, t, which is a stochastic function of true skill.

$$\log t = \log q + \mu \qquad \mu \sim N(0,\sigma_\mu^2) \qquad \mathrm{cov}(\mu,\epsilon) = 0. \tag{18}$$

The logarithmic form for the errors is chosen so that q takes on only positive values.[10] The error terms ϵ and μ correspond to random

10. This formulation allows for $q > 1$ if people receive a favorable realization of ϵ. This departs from the interpretation of q as reflecting the percentage of maximum value retained, but does not otherwise affect the analysis.

variation in workers' ability to absorb education and to translate their skill into test scores, respectively.

There are two periods. In the first, risk-neutral workers choose a level of education and obtain realizations of q, their true skill, and t, their test score. In the second, risk-neutral firms match together workers with the same test score, and pay them according to their expected productivity given their test score. Normalizing the cost of a unit of education to one, the worker's payoff V is his wage minus his education.

The analysis below follows Bulow, Geanakoplos, and Klemperer [1985] and Cooper and John [1988]. I examine only pure strategy symmetric Nash equilibria (SNE), in which all agents choose a level of education, e, which makes it optimal for each agent to choose e as his level of education. Thus, at an SNE $V_1(e,e) = 0$, where V is the payoff, the first argument is the agent's level of education, and the second argument is the level of education chosen by the other agents, who are potential coworkers.

The optimal e depends on the wage schedule, which in turn depends on the level of education chosen by all other workers in the economy, $\bar{e}$. Deriving the wage schedule requires solving a signal extraction problem to find the conditional expectation of a worker's skill given his test score and the test scores of other agents in the economy.

In equilibrium, all agents choose the effort level $\bar{e}$ and hence the expectation of $\log(q)$ and $\log(t)$ for all agents is $\log(\bar{q})$, where $\bar{q}$ is defined as $q(\bar{e})$. Firms can deduce $\bar{e}$ and thus $\bar{q}$ by observing the distribution of all agents' test scores. Since $\log(q) = \log(\bar{q}) + \epsilon$ and $\log(t) = \log(\bar{q}) + \epsilon + \mu$, and ϵ and μ are independent normals, the conditional distribution of $\log q$ for an agent with test score t given $\bar{q}$ is

$$\log q \,|\, t,\bar{q} \sim N\left(\theta \log t + (1 - \theta) \log \bar{q}, \frac{\sigma_\epsilon^2 \sigma_\mu^2}{\sigma_\epsilon^2 + \sigma_\mu^2}\right), \tag{19}$$

where θ is the share of the variance in the test score due to variance in true ability,

$$\theta \equiv \sigma_\epsilon^2/(\sigma_\epsilon^2 + \sigma_\mu^2). \tag{20}$$

Thus, if there is no testing error ($\sigma_\mu^2 = 0$), the expected skill equals the test score, whereas if there is no variation in ability to absorb education ($\sigma_\epsilon^2 = 0$), the expected skill is the average level of skill, $\bar{q}$.

Given the conditional distribution of $\log(q)$ for a single worker with test score t, a firm hiring workers of test score t has a conditional distribution of log output of

(21) $$\log(n\,\Pi_{i=1}^{n} q_i)\,|\,t,\bar{q} \sim N\left[\log n + n(\theta \log t + (1-\theta)\log \bar{q}),\, n\frac{\sigma_\epsilon^2\sigma_\mu^2}{\sigma_\epsilon^2 + \sigma_\mu^2}\right].$$

The conditional expectation of output is therefore

(22) $$E\,(n\,\Pi_{i=1}^{n} q_i)\,|\,t,\bar{q} = n\,\exp[n(\theta \log t + (1-\theta)\log \bar{q} + \log A)].$$

where A is the constant

(23) $$A \equiv \exp\left[\frac{1}{2}\frac{\sigma_\epsilon^2\sigma_\mu^2}{\sigma_\epsilon^2 + \sigma_\mu^2}\right].$$

This simplifies to

(24) $$E\,(n\,\Pi_{i=1}^{n} q_i)\,|\,t,\bar{q} = nt^{n}\bar{q}^{n\theta(1-\theta)}A^{n}.$$

By the zero profit condition, the wage is $1/n$ times the expected product:

(25) $$w(t,\bar{q}) = t^{n\theta}\bar{q}^{n(1-\theta)}A^{n}.$$

For $0 < \theta < 1$, each agent's wage is increasing not only in his own test score but also in $\bar{q}$, the skill level chosen by other agents. (Note that in the special case of no measurement error, both θ and A equal one, and the formula for the wage is the same as that derived in Section I under perfect matching.)

The marginal product of education increases with the education of other agents. The payoff V is the wage minus the cost of education:

(26) $$V(e,\bar{e}) = [g(e)\exp(\epsilon + \mu)]^{n\theta}g(\bar{e})^{n(1-\theta)}A^{n} - e.$$

Thus, for any realizations of ϵ and μ, the cross derivative of the payoff with respect to own education and others' education will be positive:

(27) $$V_{12} = n\theta[g(e)]^{n\theta-1}g'(e)[\exp(\epsilon + \mu)]^{n\theta} \times n(1-\theta)g(\bar{e})^{n(1-\theta)-1}g'(\bar{e})A^{n} > 0;$$

and hence there is strategic complementarity. Agents increase their education in response to increases in education by other agents.[11]

11. Cooper and John [1988] assume that $V_{11} < 0$. Although this game is not necessarily globally concave in own education, their analysis still applies, since the optimal choice of e must lie in a region where $V_{11} < 0$.

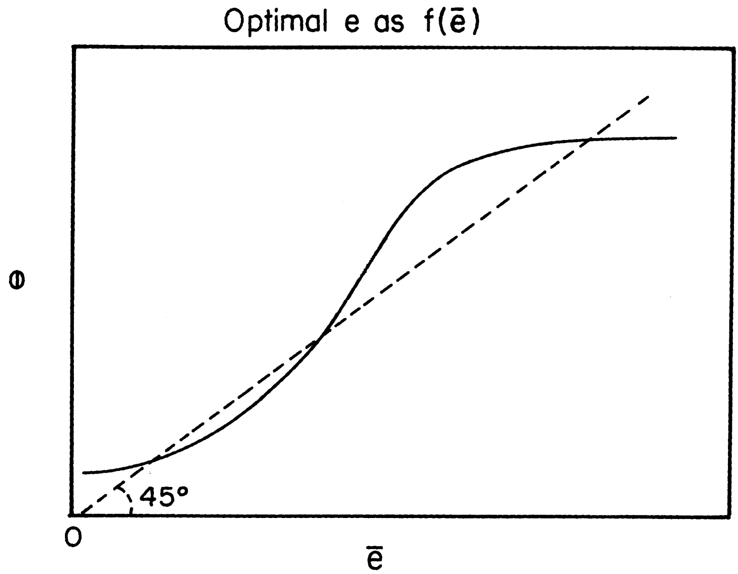

FIGURE I

Figure I shows the optimal e as a function of the level of $\bar{e}$ chosen by the other agents. Since $g(e) > 0$ and $g'(0) = \infty$, zero education can never be optimal and since $g(e)$ is bounded, the optimal e is bounded. SNE occur where the reaction function crosses the 45 degree line. Cooper and John show that a necessary condition for multiple equilibria is that the slope of the reaction function, ρ, be greater than one at some point, and a sufficient condition is that ρ be greater than one at an SNE. ρ is given by

$$\rho = -\frac{V_{12}(e,e)}{V_{11}(e,e)} = -\frac{g(e)g'(e)n(1-\theta)g'(\bar{e})}{[(n\theta - 1)g'(e)^2 + gb(e)g''(e)]g(\bar{e})}. \tag{28}$$

At an SNE, $e = \bar{e}$, so a sufficient condition for multiple equilibria is that at an SNE,

$$\rho \equiv \frac{g'(e)^2 n(1-\theta)}{(1-n\theta)g'(e)^2 - gb(e)g''(e)} > 1. \tag{29}$$

For this to hold, the denominator must be positive and smaller in absolute value than the numerator. This implies that

$$\theta < \frac{g'(e)^2 - g(e)g''(e)}{ng'(e)^2} < 1. \tag{30}$$

These inequalities are equivalent to the conditions under which $V_{11} < 0$, but $V_{11} > 0$ under perfect matching. Examination of Figure I shows that ρ must be greater than one at some SNE if there are multiple equilibria, and hence multiple equilibria are impossible if output is a globally concave function of education under perfect matching. Note also that multiple equilibria are more likely the lesser θ, the variation in true ability relative to the variation in test scores.

Since the game has positive spillovers, all SNE will be inefficient, and there will exist some level of education subsidy that improves welfare. Since there is strategic complementarity, these education subsidies will have multiplier effects. They will directly lead people to choose higher e, and this will indirectly lead people to further increase their e. Thus, small differences between countries in exogenous multiplier variables, such as tax rates, the quality of the education system, or bottlenecks, can cause large differences in q between countries. If there are multiple equilibria, variance in q between countries could be entirely endogenous. Multiple equilibria may also help explain income differences between ethnic groups within countries. If employers think an ethnic group is in a low equilibrium, they will pay a lower wage for any test score and a lower increment in the wage for any increment in the test score. Hence workers in the group will choose a lower e, validating the employers' expectations. This model of self-fulfilling statistical discrimination among microeconomically identical agents is similar to Arrow [1973] and Coate and Loury [1991], but unlike those models, which impose nonconvexity on the problem by restricting agents to one of two skill levels, qualified or unqualified, this model allows workers to take a continuum of different skill levels.

While it is not clear that this model explains a significant portion of racial discrimination in the United States, it is worth noting that historically white workers have had a higher return to education than black workers [Card and Krueger, 1992], as would be the case in the model if whites were in a higher equilibrium. Although the match with the model is far from precise, one might think that years of education completed might serve as an observable signal of how much one has learned in school, similar to the test score in the model. (Years of education would not correspond to e in the model, since e cannot be directly observed.) Under the model, legal requirements to pay black and white workers with similar observable test scores the same amount would switch both groups into the same equilibria, at least if the legal requirements were viewed as a permanent change. In fact, there is evidence that returns to education for blacks have increased since the civil rights laws of the 1960s and that blacks' education has increased accordingly [Card and Krueger, 1992]. Although Card and Krueger attribute much of the increase in returns to education for blacks to improvements in the quality of segregated black schools in the South, these improvements do not fully explain the increase, and Donahue and Heckman [1991] argue that federal policy played an important role in the 1960s. They note that this occurred despite relatively low expenditure on enforcement, as would be consistent with the existence of multiple equilibria.

III. GENERALIZATIONS AND EXTENSIONS

The matching analysis of Section I generalizes to symmetric production functions in worker skill as long as the cross derivative of output in the skill of different workers is positive. Positive cross derivatives could arise for many reasons. For example, doctors, lawyers, and academics often match with similar skill coworkers in hospitals, law firms, and universities. This may be due to learning spillovers within the firm in which high quality workers are better able to teach and learn from their coworkers.

The matching analysis and its implications thus apply to production functions that are homogeneous of degree less than one, such as

$$Y = (\Pi_{i=1}^{n} q_i)^{\psi} \qquad 0 < \psi < \frac{1}{n}. \tag{31}$$

The principal differences under production functions with decreasing returns to the skill of the workforce taken as a whole are that given differences in q create smaller rather than larger differences in output and wages; a symmetric distribution of skill leads to a distribution of income that is skewed to the left rather than to the right; and the human capital accumulation problem faced by workers is globally concave. Although strategic complementarity still arises with a decreasing returns production function under imperfect information, multiple equilibria cannot arise, since ρ can never be greater than one at an SNE if the wage is concave in e under perfect matching. This paper has concentrated on the increasing returns case, but whether decreasing or increasing returns is a more appropriate assumption is an empirical question, which presumably has different answers in different industries.

Kremer and Maskin [1993] extend the analysis to production functions with negative cross derivatives and to asymmetric production functions. Negative cross derivatives could arise, for example, if two workers were assigned to a critical task, like flying an airplane, with one serving as a backup in case the other failed to perform the task. In this case, it is optimal to match the highest and lowest skill workers together. The techniques used in Section I can be adapted to solve for equilibrium wages. Since agents match with others of different skill, each agent's wage depends on the distribution of q in the population, rather than simply on his own q. Current research focuses on endogenizing the number of workers assigned to a task, which may help bridge the gap between the efficiency units treatment of labor skill, in which quantity can simply be substituted for quality, and the O-ring approach, in which there are a fixed number of workers per task in a given production line.

A previous version of this paper (available from the author) solves for equilibrium wage schedules and assignment of workers to tasks under an asymmetric production function in which there are two types of tasks: managerial and professional tasks which are subject to multiplicative quality interaction, and unskilled tasks in which worker skill is not important. In equilibrium, agents become workers below some cutoff level of skill and managers above it. The more highly skilled managers are matched with higher quality management teams and, as in Rosen [1981] and Lucas [1978], supervise more unskilled workers. Kremer and Maskin [1993] examine a more general asymmetric production function, in which output is sensitive to the skill of all types of workers, but in different degrees. For example, the output of an orchestra might be more sensitive to the skill of the violinist than of the cellist. If workers choose their occupation before their skill is determined, the techniques of Section I can be used to solve for equilibrium wage schedules and assignment of workers to firms. The general equilibrium problem of simultaneously assigning agents to occupations and firms given their skill is more difficult. Depending on the distribution of skill, it may be optimal either for agents of similar skill to match together in firms or for agents of similar skill to take the same occupation in different firms. For example, the second highest skill musician will in some cases become a cellist with the best orchestra and in others a violinist with the second best orchestra.

In summary, the framework used in this paper readily generalizes to symmetric production functions in which quantity cannot be substituted for quality and there is a positive cross derivative in worker skill. Kremer and Maskin [1993] extend the approach to production functions with negative cross derivatives in worker skill and to asymmetric production functions.

IV. Conclusion

People in business talk about quality all the time. "Quality is Job One," "America just doesn't produce quality products anymore," "Quality Control"—all these are phrases associated with businesspeople, not economists. This paper makes a stab at modeling quality.

The paper proposes an O-ring production function in which quantity cannot be substituted for quality, shows that under this production function workers of similar skill will be matched together, and derives an equilibrium schedule of wages as a function of worker skill. Under this production function, small differences in worker skill lead to large differences in wages and output, so wage and productivity differentials between countries with different skill levels are enormous. The production function implies that workers will be sorted by quality so there will be a positive correlation among the wages of workers in different occupations within the same firm, and that firms will offer jobs to only some workers rather than paying all workers their estimated marginal product.

If tasks are performed sequentially, high-skill workers will be allocated to later stages of production. Poor countries will therefore have higher shares of primary production in GNP, and workers will be paid more in industries with high value inputs. If firms can choose among technologies with different numbers of tasks, the highest skill workers will use the highest n technology. This is consistent with the tendency of rich countries to specialize in complicated products, and, given a correlation between n and firm size, with the larger average firm size in rich countries and the positive correlation between firm size and wages within countries. These predictions of the model match stylized facts about the world, and although each of these facts may be due to a variety of causes, together they suggest that O-ring production functions are empirically relevant.

Imperfect matching of workers due to imperfect information about worker skill leads to positive spillovers and strategic complementarity in investment in human capital. Thus, subsidies to investment in human capital may be Pareto optimal. Small differences between countries in such subsidies or in exogenous factors such as geography or the quality of the educational system lead to multiplier effects that create large differences in worker skill. If strategic complementarity is sufficiently strong, microeconomically identical nations or groups within nations could settle into equilibria with different levels of human capital.

The matching results and their implications apply to a general symmetric production function in which quantity cannot be substituted for quality, as long as there is a positive cross derivative in worker skill. Current research focuses on adapting these techniques to solve for equilibrium wages and assignment of workers to firms under production functions with negative cross derivatives and under asymmetric production functions.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

REFERENCES

Arrow, K., "The Theory of Discrimination," in O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets* (Princeton, NJ: Princeton University Press, 1973).
Barro, Robert, "Economic Growth in a Cross Section of Countries," *Quarterly Journal of Economics,* CVI (May 1991), 407–43.
Becker, Gary, *Treatise on the Family* (Cambridge: Harvard University Press, 1981).
Brown, Charles, and James Medoff, "The Employer Size-Wage Effect," *Journal of Political Economy,* XCVII (1989), 1027–59.
Bulow, Jeremy I., John D. Geanakoplos, and Paul D. Klemperer, "Multimarket Oligopoly: Strategic Substitutes and Complements," *Journal of Political Economy,* XCIII (1985), 488–511.
Card, David, and Alan Krueger, "School Quality and Black-White Relative Earnings: A Direct Assessment," *Quarterly Journal of Economics,* CVII (1992), 151–200.
Clague, Christopher, "Relative Efficiency, Self-Containment, and Comparative Costs of Less-Developed Countries," *Economic Development and Cultural Change* XXXIX (1991a), 507–530.
____, "Factor Proportions, Relative Efficiency and Developing Countries Trade," *Journal of Development Economics,* XXXV (1991b), 357–80.
Clark, Gregory, "Why isn't the Whole World Developed? Lessons from the Cotton Mills," *Journal of Economic History,* XLVII (March 1987), 141–73.
Coate, Stephen, and Glenn Loury, "Affirmative Action as a Remedy for Statistical Discrimination," mimeo, Harvard University, 1991.
Cooper, Russell, and Andrew John, "Coordinating Coordination Failures in Keynesian Models," *Quarterly Journal of Economics,* CIII (August 1988), 441–64.
Dechert, W. Davis, and Kazuo Nishimura, "A Complete Characterization of Optimal Growth Paths in an Aggregated Model with a Non-Concave Production Function," *Journal of Economic Theory,* XXXI (1983), 332–54.
Donahue, John, and James Heckman, "Continuous Versus Episodic Change: The Impact of Affirmative Action and Civil Rights Policy on the Economic Status of Blacks," *Journal of Economic Literature,* XXIX (1991), 1603–43.
Galor, Oded, and Joseph Zeira, "Income Distribution and Macroeconomics," Working Paper No. 89-25, 1989.

Jovanovic, Boyan, "Job Matching and the Theory of Turnover," *Journal of Political Economy,* LXXXVII (1979), 972–90.
Katz, Lawrence, and Lawrence Summers, "Industry Rents: Evidence and Implications," *Brookings Papers on Economic Activity: Microeconomics* (1989), 209–75.
Kremer, Michael, and Eric Maskin, "Matching Heterogeneous Workers to Firms and Occupations: Preliminary Notes," 1993.
Kwon, Goohoon, "A Comparison of Sectoral Productivity Levels of the USSR and USA," mimeo, 1992.
Lucas, Robert, "On the Size Distribution of Business Firms," *Bell Journal of Economics,* IX (1978), 508–23.
——, "Why Doesn't Capital Flow from Rich to Poor Countries?" *American Economic Review,* LXXX (May 1990), 92–96.
Mankiw, N. Gregory, David Romer, and David N. Weil, "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics,* CVII (May 1992), 407–38.
Miller, Frederick, "The Distribution of Ability and Earnings Once Again," mimeograph, Johns Hopkins, 1983.
Murphy, Kevin, Andrei Shleifer, and Robert Vishny, "The Allocation of Talent: Implications for Growth," *Quarterly Journal of Economics,* CVI (1991), 503–30.
Rosen, Sherwin, "The Economics of Superstars," *American Economic Review,* LXXI (1981), 845–58.
——, "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics,* XIII (Autumn 1982), 311–23.
Sah, Raaj, and Joseph Stiglitz, "Human Fallibility and Economic Organization," *American Economic Review,* LXXV (May 1985), 292–97.
Sah, Raaj, and Joseph Stiglitz, "The Architecture of Economic Systems: Hierarchies and Polyarchies," *American Economic Review,* LXXVI (September 1986), 716–27.
Sattinger, Michael, "Comparative Advantage and the Distributions of Earnings and Abilities," *Econometrica,* XLIII (1975), 455–68.
Sobel, Joel, "How to Count to One Thousand," *Economic Journal,* CII (January 1992), 1–8.
World Bank, *World Development Report 1990* (New York: Oxford University Press, 1990).