6-1-1984

Job Discrimination, Market Forces and the Invisibility Hypothesis

Paul R. Milgrom

Sharon Oster

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 708

JOB DISCRIMINATION, MARKET FORCES AND THE INVISIBILITY HYPOTHESIS

Paul Milgrom

and

Sharon Oster

June, 1984

JOB DISCRIMINATION, MARKET FORCES
AND THE INVISIBILITY PROCESS

Paul Milgrom

School of Organization and Management, Department of Economics
and Cowles Foundation, Yale University

and

Sharon Oster

School of Organization and Management, Yale University

ABSTRACT

The Invisibility Hypothesis holds that the job skills of disadvantaged workers are not easily observed by potential new employers, but that promotion enhances visibility and alleviates this problem. Then, at a competitive labor market equilibrium, disadvantaged workers will be paid less on average and promoted less often than other workers with the same education and ability, even if their employers are unprejudiced and know their workers' abilities. As a result of the discriminatory wage and promotion policies, disadvantaged workers will experience lower returns to investments in human capital than other workers. An affirmative action program that eliminates discrimination and brings about efficiency initially forces the promotion of unqualified workers.

Key Words: Discrimination, Human Capital, Job Discrimination, Labor Markets, Visibility.

Journal of Economic Literature Classifications:
Primary - 821
Secondary - 026, 826

Acknowledgments. We would like to thank Rick Levin and the participants of the Yale Microeconomics Seminar and the Cowles seminar for their helpful comments. Milgrom's work was partially supported by NSF grant IST-8411595 to Yale.

JOB DISCRIMINATION, MARKET FORCES AND THE INVISIBILITY HYPOTHESIS

by Paul Milgrom and Sharon Oster

Among non-economists, job discrimination is rarely regarded as a phenomenon that can be understood in purely economic terms, without reference to its rich sociological context. Indeed, it is more often seen as class phenomenon, part of a tightly woven, self-reinforcing fabric of discrimination, which encompasses housing, education, jobs, club memberships, and more.

In contrast, classical economic theory holds that a worker's pay and chances for promotion depend only on his or her productive abilities. The presence of a few prejudiced employers who discriminate against, say, non-white workers or women is irrelevant according to this view: the remaining employers will be led by the pursuit of profits to hire the underpaid and promote the underemployed. Competition among unprejudiced employers will result in rising wages for victims of discrimination. Therefore, the existence of job discrimination is at worst a minor and temporary phenomenon, or so the usual argument goes.

If one acknowledges that there is discrimination, which we take to mean that some employers knowingly underpay and refuse to promote talented non-whites and women, then the puzzle is: How can market forces allow such a state of affairs to continue? One possibility, explored by Becker (1957), is that individuals have a taste for discrimination -- a widespread willingness to earn lower profits or receive lower wages in order to maintain segregation in the workplace. That view does not deny market forces, but holds that if discrimination persists in the labor market, then, while terribly inequitable, it must nevertheless be Pareto efficient.

Another view, advanced by Phelps (1972), explains discrimination without resorting to an assumption of prejudice in the workplace. Phelps's theory is based on the hypothesis that traditional indicators of ability are less informative when applied to disadvantaged workers. Rational employers therefore place less reliance on them. As a result, a kind of discrimination occurs, since employers evaluate indicators of ability differently for members of different groups. However, this "statistical theory of discrimination" does not by itself explain lower average wages for disadvantaged workers; it predicts that the average wage paid to the workers in any group that is distinguished by an observable characteristic (such as sex or race) depends only on those workers' productive prowess.

A recent, important paper by Lundberg and Startz (1983) extends the statistical theory of discrimination to include a human capital investment choice by the workers. Like Phelps, Lundberg and Startz assume that each worker is paid his or her expected marginal product, given the available indicators. Then, since disadvantaged workers' job skills are less accurately tested than those of other workers, they have less incentive to make unobservable investments in human capital.[1] Therefore, they acquire less human capital, and their average productivity and wages are correspondingly lower. This theory, however, does not consider issues of job assignment[2], and does not account for employers who knowingly discriminate in wages and promotions against talented workers from disadvantaged groups.

It is our view that job discrimination is just one part of the larger fabric of discrimination. It seems clear that job discrimination, with its attendant low incomes and poor job opportunities, aggravates the problems of housing, club memberships, low self-esteem, etc., that plague disadvantaged workers. The effect we wish to study is the one running in the other direction: How might impediments created by non-market aspects of discrimination twist market outcomes to the disadvantage of minorities and women and workers from impoverished backgrounds?

The focus of our explanation of job discrimination is the differential visibility of the job skills of workers with different backgrounds. First, we hypothesize that although the skills of any worker are known perfectly to his or her own employer, a disadvantaged worker is less able to make those skills visible to other potential employers. The worker is therefore less able to generate the wage offers necessary to command a wage equal to his or her marginal revenue product.[3] Second, we hypothesize that the visibility problem is greatest for workers in the lowest level jobs, and is alleviated by being assigned to a higher level job. We combine these two hypotheses into a single one, which we call the Invisibility Hypothesis.

The Invisibility Hypothesis reflects a variety of factors both inside and outside the workplace that are too rarely considered in economic theories. Simple prejudice -- in the form of misperceptions rather than antipathy -- can cause an employer to overlook a potentially good employee. So can the failure of an employee to "toot his own horn," whether the reluctance to do so comes from shyness or pride or cultural taboos. The existence of clubs which limit the membership of women, non-whites, or religious or ethnic minorities; job segregation which is not *per se* inefficient but which keeps some people out of view; exclusive neighborhoods; out-of-town conventions that are hard for some working mothers to attend -- all of these things contribute to a separation that makes some workers less visible to potential new employers. This invisibility leads to wage distortions by impairing the competition among firms for workers. Using the Invisibility Hypothesis, we are able to trace the effect that misperceptions and segregation have on the employment practices of profit-seeking firms.

We investigate the implications of the Invisibility Hypothesis in a competitive market equilibrium model in which there are only two types of jobs, low level and high level. We assume that more talented workers are more productive in any kind of job, and have a comparative advantage in higher level jobs. Also, initially, we assume that employment contracts consist simply of a wage offer and a guarantee against wage cuts or layoffs[4]. In our model, employers buy labor at market-clearing prices to maximize profits, and workers invest in education and other forms of human capital to maximize expected wages net of investment costs. We assume that there are two kinds of workers: Visibles, whose abilities are known to all potential employers, and Invisibles, whose abilities are known only to their own current employers. Among the conclusions derived from our model are the following:

1. Average wages will be lower for Invisibles than for Visibles. (Hence, groups such as women and certain minorities which have a relatively high proportion of Invisibles among their members will have lower average wages.)

2. Wage variability will be less for Invisibles, both in total and in each kind of job.

3. Promotions will be less probable for Invisibles. The ablest Invisibles will be inefficiently assigned to low level jobs.

4. Conditional on being promoted, wages will be lower for Invisibles. No Invisibles will be in the most highly paid group of workers.

5. The returns to unobservable investments in human capital will be lower for Invisibles.

6. The returns to investments in general education[5] are lower for Invisibles provided that either of the following two conditions hold: (i) workers of below average ability never have a comparative advantage at high level jobs or (ii) investments in general education are predominantly most productive for the ablest workers.

7. As a corollary to 5 and 6, an Invisible will acquire less general education and less unobservable human capital than an equally able Visible when the stated conditions hold.

8. Invisibles whose abilities are most enhanced by education will sometimes forgo education, while others whose skills benefit less acquire more education.

9. If workers know their own abilities, then an education subsidy that leads Invisibles to acquire as much education as Visibles will still lead to an inefficient distribution of education among members of the Invisible group. That, together with the job discrimination, leads to persistently lower wages and lower probabilities of promotion for Invisibles.

10. An affirmative action program that sets quotas on the number of Invisibles promoted is of limited value, since the wrong employees will be promoted.

11. An affirmative action policy that sets no quotas but forces wage parity between Visibles and Invisibles in each job class (low level and high level) will result in increased job segregation (lower promotion rates for Invisibles and higher rates for Visibles) and lower average wages for Invisibles.

12. An affirmative action program that sets both quotas and wage standards will, in the short run, have ambiguous effects on efficiency. It will lead to the promotion of some qualified workers who had been inefficiently assigned to low level jobs, but it will also result in the promotion of some unqualified Invisibles, who will be overpaid relative to their productivity. In the long run, if the quotas and wages are set correctly, such an affirmative action program will improve efficiency of both job assignments and human capital acquisition decisions.

13. Unless the underlying invisibility problem is solved, an affirmative action program that is temporary cannot eliminate discrimination permanently.

Because our model includes all of the usual labor market forces contained in the neoclassical model and still leads to discrimination, it establishes that the effects of segregation and prejudice, as embodied in the Invisibility Hypothesis, can reduce -- and even twist -- the propensity of market forces to eliminate discrimination. Moreover, although our theory identifies employers with firms, the analysis need not stop there: invisibility can be a problem even within a corporation. In an analysis closely akin to ours, Kanter (1977) has identified a phenomenon that she calls the "dilemma of indispensability," in which bosses try to conceal able secretaries from others in the same firm who might bid them away. Secretaries are trapped in lower level jobs by their own talents -- their own indispensability to their bosses, and by their invisibility. This sort of discrimination by individual managers with few employees is hard to detect, because statistical tests of discrimination with small samples tend to be unreliable. Partly for that reason, policy measures aimed at eliminating discrimination by individual managers are relatively ineffective.

THE MODEL

We analyze a simple model, in which all firms are identical and have production functions $f(x,y)$ with two inputs: labor in low level jobs x and labor in high level jobs y. Employment contracts specify a wage $w(e,\theta)$ for those with education e and known ability θ and a wage $w(e)$ for those of unknown ability. Wage reductions and layoffs are prohibited. Workers of known ability consist of the Visibles plus any Invisibles who have been promoted.

The two labor inputs, x and y, are measured in efficiency units. A worker of native ability θ who has acquired education e provides $\alpha(e,\theta)$ efficiency units of low level labor or $\alpha(e,\theta)\ \beta(e,\theta)$ units of high level labor, depending on his or her job assignment. Note that β is a measure of the worker's comparative advantage in high level jobs. It is assumed that abler workers are more productive in both kinds of jobs and have a comparative advantage in high level jobs. Thus, θ is a real variable and both α and β are increasing in θ.

Our technical assumptions are these: The worker productivity functions α and β are positive, bounded, and continuously differentiable with $\partial\alpha/\partial\theta > 0$ and $\partial\beta/\partial\theta > 0$. The production function f is a smooth neoclassical production function: it is twice continuously differentiable with constant returns to scale and a diminishing marginal rate of substitution among its two inputs. Positive amounts of both kinds of inputs are required for production. To keep notation to a minimum, f is measured in revenue units.

The numbers N_V of Visibles and N_I of Invisibles per firm are specified exogenously. Also, at the time wages are determined, the human capital acquisition decisions are assumed to have already been made, so the distribution of education and talents among workers are parameters from the firms' point of view. Let the joint density of education and ability among Visibles be $g(e,\theta)$ and among Invisibles be $h(e,\theta)$. Finally, we assume that among those of unknown ability, the firm hires a perfect cross-section, so that the distribution of abilities given education among its workers is the same as in the population at large.

The firm's problem is to choose how many workers of each type to hire, and how to assign workers to jobs. Let $n(e,\theta)$ designate the number of Visibles with education e and ability θ that the firm hires, and let m(e) be the number of Invisibles hired with education e. Let $p(e,\theta)$ and $q(e,\theta)$ be the proportions of Visibles and Invisibles, respectively, with education e and ability θ that are assigned to high level jobs. The functions n, m, p, and q are the employer's choice variables. These choices determine the quantities of efficiency labor $x = x_V + x_I$ and $y = y_V + y_I$ and the total wage bill $W = W_V + W_L + W_H$ by the equations:

$$x_V = \iint \alpha(e,\theta)\ n(e,\theta)\ [1 - p(e,\theta)]\ d\theta\ de \quad (1a)$$

$$x_I = \iint \alpha(e,\theta)\ m(e)\ h(\theta|e)\ [1 - q(e,\theta)]\ d\theta\ de \quad (1b)$$

$$y_V = \iint \alpha(e,\theta)\ \beta(e,\theta)\ n(e,\theta)\ p(e,\theta)\ d\theta\ de \quad (1c)$$

$$y_I = \iint \alpha(e,\theta)\ \beta(e,\theta)\ m(e)\ h(\theta|e)\ q(e,\theta)\ d\theta\ de \quad (1d)$$

$$W_V = \iint w(e,\theta)\ n(e,\theta)\ d\theta\ de \quad (2a)$$

$$W_L = \iint \underline{w}(e)\ m(e)\ h(\theta|e)\ [1 - q(e,\theta)]\ d\theta\ de \quad (2b)$$

$$W_H = \iint \max\{\underline{w}(e), w(e,\theta)\}\ m(e)\ h(\theta|e)\ q(e,\theta)\ d\theta\ de. \quad (2c)$$

The labor inputs for the firm consist of the low level efficiency units x and high level units y provided by Invisibles and Visibles, as determined by the hiring and job assignment policies. Visibles must be paid their market wage, regardless of the job assignment. Invisibles assigned to low level jobs are paid their entry wages. If an Invisible is assigned to a high level job and if his or her market wage exceeds the entry wage, then the higher market wage must be paid to prevent the worker from being bid away. All of these facts are reflected in the set of equations above.

The labor market is in short-run equilibrium when, at specified wages, the number of Visibles demanded at each education-and-ability combination and the optimal number of Invisibles demanded at each education level are equal to their respective supplies. In formal terms, these market-clearing conditions take the form:

$$n(e,\theta) = N_V\ g(e,\theta) \qquad (3a)$$

$$m(e) = N_I\ h(e)\ . \qquad (3b)$$

Proposition 1. At equilibrium, the wages and assignments of the workers are given by:

$$w(e,\theta) = \alpha(e,\theta)\ \max(f_1,\ f_2\ \beta(e,\theta)) \qquad (4)$$

$$\underline{w}(e) = f_1 \frac{\int \alpha(e,\theta)\ h(\theta|e)\ [1 - q(e,\theta)]\ d\theta}{\int h(\theta|e)\ [1 - q(e,\theta)]\ d\theta} \qquad (5)$$

$$p(e,\theta) = \begin{cases} 1 & \text{if } \theta > \theta^*(e) \\ 0 & \text{if } \theta < \theta^*(e) \end{cases} \qquad (6)$$

$$q(e,\theta) = \begin{cases} 1 & \text{if } \theta^*(e) < \theta < \theta^{**}(e) \\ 0 & \text{otherwise} \end{cases} \qquad (7)$$

with $f_1 = \partial f/\partial x$, $f_2 = \partial f/\partial y$. The breakpoint $\theta^*(e)$ is given by equation (8a) if a solution exists and generally by (8b), and $\theta^{**}(e)$ is given by (9):

$$\beta(e,\theta^*(e)) = f_1/f_2 \qquad (8a)$$

$$(\beta(e,\theta) - f_1/f_2)\ (\theta - \theta^*(e)) \geq 0 \qquad (8b)$$

$$\alpha(e,\theta^{**}(e)) = \underline{w}(e)/f_1 \qquad (9)$$

The partial derivatives in Proposition 1 are evaluated at the arguments $x = x_V + x_I$ and $y = y_V + y_I$ determined from equations (1a)-(1d). Given the numerical values of f_1 and f_2, the other variables can be computed as follows. First, $\theta^*(e)$ and $w(e,\theta)$ are determined from (8) and (4) for all educational attainments e.[6] Then, $\theta^{**}(e)$ and $\underline{w}(e)$ are determined by equations (5), (7) and (9), as shown in Figure 1. Notice that, in view of the promotion policy described by (7), equation (5) expresses the average productivity variable $\underline{w}$ is as a quasi-concave function of θ^{**}. As θ^{**} is raised, workers of below average productivity may first be promoted out of low level jobs so the average productivity $\underline{w}$ rises. However, eventually, raising θ^{**} further causes workers of above average productivity to be promoted, so the average productivity falls. Equation (9) then asserts that the marginal worker excluded by an increase in θ^{**} is of average ability for that group. This implies that the equilibrium choice θ^{**} is the choice that maximizes $\underline{w}$ in (5).[7]

The essential content of Proposition 1 is seen from a comparison of (6) and (7): Invisibles suffer discrimination in promotions. Visibles assigned to higher level jobs are precisely those who are more productive in those jobs, that is, those for whom $\theta > \theta^{*}(e)$. Invisibles assigned to high level jobs must be both more productive in those jobs <u>and</u> less able than an average low level Invisible worker with the same education. There may well be <u>no</u> such Invisibles, and in that case the Proposition asserts that none are assigned to high level jobs.

Why do firms practice this form of discrimination in which the ablest Invisibles are not promoted? Employers earn rents on the ablest Invisibles in low level jobs, since low level Invisibles are paid the <u>average</u> marginal product of all workers in their class, while the ablest individuals, those with $\theta > \theta^{**}(e)$, have marginal products in excess of the class average. Since the ability of a promoted worker becomes known to other potential employers, it is impossible for a firm to earn rents on a high level worker. A profit-maximizing employer will inefficiently assign the ablest Invisibles to low level jobs, since to do otherwise would jeopardize the rents earned on those individuals. In contrast, since the wages of Visibles are fixed independently of their job assignments, it is profitable for the employer to assign Visibles to jobs in an efficient manner. In a competitive labor market, employers earn zero excess profits on any distinguishable class of workers, so any efficiency losses must be directly reflected in the relative wages of the two groups. We summarize these conclusions in the form of a Proposition.

Proposition 2. Suppose that the joint distributions of education and ability were the same for Visibles and Invisibles. Then the average wage earned by Invisibles would be lower than for Visibles. Moreover, the probability of promotion (that is, of assignment to a high level job) would be lower and, if any Invisibles were promoted, the average wage among those promoted would be lower than for Visibles. The highest wage paid to a Visible would be higher than the highest wage paid to any Invisible.

Now consider the difference in the wages paid to Visibles and Invisibles as a function of their skill levels. It is entirely possible in this model for $\theta^{**}(e)$ to be less than $\theta^{*}(e)$ for some or all values of e, in which case all Invisibles who are able enough to be promoted are also a source of rents to their employers. In this case, no Invisible workers will be promoted and all will be paid the same entry wage. Then, the difference between the wages paid to Visibles of type (e,θ) and Invisibles of the same type is $w(e,\theta) - \underline{w}(e)$. That difference is increasing in θ because abler Visible workers command higher wages (as is readily verified from Proposition 1), while Invisibles all earn the same wage.

The other possibility is that $\theta^{**}(e)$ exceeds $\theta^{*}(e)$; that is, some Invisibles (albeit not the most able) are promoted and earn a higher wage. In that case, the difference in wages is still nondecreasing in θ: it is equal to zero for θ between $\theta^{*}(e)$ and $\theta^{**}(e)$ and to $w(e,\theta) - \underline{w}(e)$ outside that range. Figure 2 illustrates this case.

The observation that the difference between the wage of a Visible and an Invisible of equal ability θ is increasing in θ has important implications for the workers' incentives to acquire productive ability. Thus, suppose that a worker can accumulate human capital that enhances his or her ability but which is not directly observed by potential employers. Let ε denote the worker's expenditure on such human capital, and suppose that larger values of ε shift the distribution of θ up stochastically[8] in a way that is subject to initially large but eventually decreasing returns. Thus, letting $F(\theta|\varepsilon)$ be the distribution of ability given ε, we suppose that F always has a corresponding density, that F is decreasing and strictly convex in ε, that the support of F does not depend on ε, and that $\partial F/\partial\varepsilon$ evaluated at $\varepsilon = 0$ is $-\infty$. Suppose, too, that workers invest in human capital to maximize expected wages net of investment costs. In the case where $\theta^{**}(e) < \theta^{*}(e)$, an Invisible's wage never depends on ability, and it is clear that the Invisible will set $\varepsilon = 0$, while the Visibles will make positive unobservable investments in human capital.

For the case $\theta^*(e) < \theta^{**}(e)$, the difference in (Visible minus Invisible) wages as a function of ability is still nondecreasing so if Invisibles do not know their own abilities, the marginal returns to unobserved investments in human capital are higher for Visibles in that case, too.[9] Next, suppose Invisibles do know their own native abilities μ; for example let $\theta = \mu + \phi(\varepsilon)$. Then the marginal return to ε is sometimes less, and never more, for Invisibles than for Visibles. The marginal returns are always lower for the ablest Invisibles. Therefore:

<u>Proposition 3</u>. Visibles invest at least as much in unobserved human capital as Invisibles with the same ability and education, regardless of whether workers know their own abilities at the time the investment is made. If workers do know their own abilities, then the ablest Invisibles always invest too little.

The foregoing analysis was carried out under the assumption that human capital is a one-dimensional variable. In the economy we have modelled, there are different kinds of jobs, and it therefore makes sense to assume that there are different kinds of human capital. In such a world, discrimination distorts not only the level of human capital investment, but also the kind of human capital acquired.

Consider the problem facing an Invisible, holding the observable education variable e fixed. Let us now regard the productivity parameters $\alpha(\varepsilon_1, \theta)$ and $\beta(\varepsilon_2, \theta)$ as functions of the unobserved choice $\varepsilon = (\varepsilon_1, \varepsilon_2)$. The logic of the labor market equilibrium is unaffected by this change: It is still profitable for employers to hold back Invisibles with high productivity in low level jobs, because they are a source of rents, whether or not they have a comparative advantage in high level jobs.

There are two cases to be considered. If $\theta^{**}(e) < \theta^{*}(e)$, then Invisibles with education e are never assigned to high level jobs and their pay in low level jobs is independent of the human capital acquired. In such a case, the question of investing in the wrong kind of human capital does not arise; Invisibles invest zero in both kinds.

The second possibility is that $\theta^{**}(e) > \theta^{*}(e)$, so that some Invisibles, but not the ablest ones, are promoted. In that case, if Invisibles do not know their own ability levels θ, there is too little incentive for an Invisible to invest in the job skills that give a comparative advantage in high level jobs. Moreover, since Invisible workers are assigned inefficiently to jobs, there is too little incentive to invest in generalized human capital as well. The actual optimal choices for the two types of workers depends on the technology of acquiring human capital. Thus;

Proposition 4. Suppose that $\partial\alpha / \partial\varepsilon_1$ and $\partial\beta / \partial\varepsilon_2$ are positive, declining, and infinite at zero. Let $C(\varepsilon_1, \varepsilon_2)$ be the cost of acquiring unobservable human capital and suppose that it is increasing, differentiable and convex and that $\partial^2 C / \partial\varepsilon_1 \partial\varepsilon_2 \leq 0$. Then Invisibles who make their human capital acquisition decisions without knowing their abilities set both ε_1 and ε_2 below their efficient levels.

For workers who know their own abilities, however, a different conclusion emerges. The ablest workers, hoping to be assigned to high level (high paying) jobs, seek to keep α small, so that they will have below average productivity in low level jobs, and will not be a source of rents to their employers.[10] Depending on the technology of skill acquisition, these workers may then acquire either more or less of the skills that enhance comparative advantage.

Although a strategy of avoiding skills that give one a comparative advantage in low level jobs may be effective for the worker in this very imperfect world, it is, of course, inefficient from society's point of view. First, to the extent that these skills contribute not only to productivity in low level jobs, but to productivity in high level jobs as well, they will be acquired at inefficiently low levels. Second, despite this strategy, some of the able Invisibles will be assigned to low level jobs, and they will have acquired inefficiently low skill levels given their eventual assignments. Thus, the ablest Invisibles make distorted, inefficient choices; they acquire too little of the skills which weaken their comparative advantage in high level jobs.

It is unreasonable to suppose that human capital acquisition is entirely unobserved, and it was for that reason that we included the observable education choice in our model of productivity. Unlike the choice of unobservable human capital ε, observable education e which raises a worker's productivity necessarily raises his or her wage. In general, in our framework, Visibles are paid their marginal products, while Invisibles are paid their expected marginal products minus the efficiency loss that comes from the failure to promote the ablest of them. Thus, the marginal return to observable investments in education will higher for Invisibles than for Visibles if more educated Invisibles suffer less discrimination (have less lost output due to inefficient assignments) than less educated ones. Conversely, if more educated Invisibles suffer more discrimination, then the marginal return is lower for Invisibles. To determine which of these cases applies, one must specify how investments in education affect productivity in each kind of job.

We can identify two cases in which the marginal return to observable investments education is lower for Invisibles. First, if the sole effect of education were to enhance the productivity of the ablest workers in high level jobs, then since the ablest Invisibles are never assigned to high level jobs, the return to education for Invisibles would be zero. The second is the case in which workers of average productivity in low level jobs can never have a comparative advantage in high level jobs (so that

$\theta^*(e) > \theta^{**}(e)$ for all e) and in which education enhances a worker's comparative advantage in high level work (so that $\partial\beta/\partial e > 0$). Then more educated workers suffer more discrimination (the efficiency loss from inappropriate assignments is greater for the better educated). In this case again, the marginal return to education is lower for the Invisibles.

Evidently, education is more valuable to an Invisible than to a Visible if (i) it enhances productivity in the low level job without enhancing productivity in the high level job, or (ii) it raises the productivity of low-skilled workers in low level jobs, without equally raising the productivity of medium- and high-skilled workers in low level jobs. In both of these cases, education reduces the efficiency loss associated with improper job assignments at the market equilibrium.

It is interesting to interpret these findings in terms of the kinds of education that members of disadvantaged groups might seek to acquire. Invisibles who study a trade are likely to be more efficiently assigned, according to our theory, than those who acquire a general education: Potential employers can identify that an electrician assigned to mow lawns is underemployed, but cannot make the same determination for a clerk who would be more efficiently used as an office manager. Similarly, among those with college educations, those who study subjects like engineering acquire specific skills that make it easier to identify an underemployed worker. Such studies ought to have higher returns for Invisibles, according to our theory, than a general education, even if the latter actually contributes more to the worker's productive capabilities.

PRIVATE AND PUBLIC STRATEGIES FOR REDUCING DISCRIMINATION

As a result of their invisibility, there is a class of workers who are subject to wage discrimination. This discrimination brings with it a real, and potentially large, efficiency loss. The efficiency loss takes two forms in our theory. First, the most talented Invisibles will be inefficiently assigned to low level jobs. Secondly, Invisibles in general will acquire too little human capital and, moreover, they will acquire the wrong sorts of human capital.

Given that discrimination exists we want next to consider whether there are viable strategies that either individual workers or government could use to reduce or eliminate this discrimination. We also want to consider whether there exist any private labor contracts which firms might make with members of disadvantaged groups which would both reduce discrimination and yield profits for the firm. If there are any such contracts, one might expect firms to offer them, and then there would be a force at work in the market to eliminate the discriminatory labor market equilibrium that we have described.

Let us assume, for the reasons discussed earlier, that there is a disproportionate number of women and minorities among the Invisible group. Suppose that, in an attempt to remedy discrimination, the government imposes a quota on a firm, which takes the form that a certain percentage of the women and minorities in its labor force must be promoted. If the program is to have any chance of being effective, the quota must be binding. Let us assume that it is binding: p under the quota is in excess of the optimal p from the firm's point of view.

Assume first that $\theta^{**}(e) > \theta^{*}(e)$. In this case, the firm has three instruments available to meet the quota, which we organize as follows.

1. The firm can increase the number of Visible women and minorities promoted. This would require firms to promote people with skill levels $\theta < \theta^{*}$, and would create an efficiency loss both for the firm and for society as a whole.

2. The firm can increase the number of Invisible women minorities promoted. This can be done either by:

(a) Promoting a larger group of the Invisibles with low skill levels; that is, those with $\theta < \theta^{*}$. This also creates an efficiency loss for both the firm and the society.

(b) Promoting a group of the women (minorities), on whom the firm is currently earning a rent, $\theta > \theta^{**}$. This creates a loss for the firm equal to the rent lost, but involves an efficiency gain from the point of view of society as a whole, since a larger fraction of talented women and minorities will be promoted.

The firm will use a cost-minimizing mix of these three instruments to meet the quota, but only option 2(b) is efficiency enhancing. First order conditions tell us that, in general, firms will use a mix of the three instruments, with the mix depending on the distribution of e and θ as well as on the functions α and β. One can show that, depending on these parameters, the quota may either increase or decrease (but can never eliminate) the costs of inefficient job assignments.

The second case involves $\theta^{*}(e) > \theta^{**}(e)$, so that the firm earns rents on all Invisible women and minorities of sufficient skill levels to promote. In that case, the marginal cost to the firm of promoting unqualified women and minorities from the Visible group is zero, since that group contains some marginally unpromoted workers. The marginal cost of promoting Invisible women and minorities is positive, since $\theta^{*}(e) > \theta^{**}(e)$. Thus, any small quota will unambiguously increase efficiency losses without improving the wages or opportunities of the subgroups of Invisible women and minorities who have suffered from discrimination.

If the quota program does not specify that the women and minorities assigned to high level jobs be promoted from within, there is yet another option open to the firm: It can try to hire qualified (Visible) women and minorities from other firms. If the quotas affect only a few firms, this option will be the cheapest one to implement, and it will result in no change in the status of any Invisible worker and no enhancement of efficiency. In such a case, the quota system is entirely ineffective.

In our theory, a quota system cannot solve the discrimination problem. It is important to distinguish our reasoning from that of economists such as Welch (1981). In his model, firms do their best to assign workers efficiently to jobs, so that any quota causes an inefficient misallocation of jobs between majority and minority workers. Within the minority category, there is efficient assignment; that is, the quota leads to promotion for higher quality workers. In our model, the quota may increase inefficiency by exacerbating the problem of mismatching of workers to jobs within the pool of women and minority workers.

With only slight modifications, the same analysis applies to a firm's private offer of a contract with a specified promotion rate. Given the firm incentives described above it will not be credible that a firm would apply its fixed promotion rate to the most productive of the disadvantaged workers.

A second possibility is for the government to specify the wage schedule to be offered to the disadvantaged workers. A first variation, investigated by Lundberg and Startz (1983), is to require that firms offer all managerial employees the same wage schedule, $w(\theta)$. In our model, this policy has no effect; firms already follow this practice because the observed variable is a perfect indicator of productivity. The focus of our theory is the inefficient assignment of workers to jobs, and the inefficient acquisition of human capital. The equal wage schedule has no effect on either of these problems.

A third, more radical option, is for the government to require employers to pay the same average wages to each identified class of workers (men, women, minorities, etc.) in the same job category (in our model there are two categories, low level and high level). The effect of such a policy is to make it more costly for a firm to promote an Invisible whose skills justify a below average wage. Holding the characteristics of the labor force fixed and tracing the effects of the policy on the labor market equilibrium, it is straightforward to verify (i) that job segregation is increased -- no Invisibles are promoted but that more Visibles are promoted than in the original equilibrium, (ii) that job assignments become less efficient, (iii) that the average skill level of unpromoted Invisibles falls, (iv) that the marginal product of low level labor falls while the marginal product of high level labor rises, and (v) consequently that the wage earned by Invisibles falls while those earned by promoted Visibles rises. Thus, such a policy reduces the wages and promotion opportunities of Invisibles further and exacerbates the problem of inefficient assignments.

A final proposal is to set both quotas for hiring and promotion of members of disadvantaged groups and a requirement that the average wage paid to women and minorities in high level jobs be the same as for the advantaged group. Such a policy would alter the incentives of individual firms in a complicated way. For example, it may pay a firm to hire a Visible woman executive at a wage exceeding her productivity in order bring up average wages for women managers and permit continued discrimination against Invisible women. If the affirmative action program were narrowly applied, for example only to large government contractors, firms might achieve compliance by hiring and promoting able Visibles from the disadvantaged groups, a policy which leaves the wage paid to each worker unchanged and so carries no benefits for any disadvantaged workers. If, however, the program were applied broadly, the labor market equilibrium would be upset.

It is not difficult to trace the effects of a broad application of a quota-and-wage-scale policy through to equilibrium. For any fixed distribution of abilities among Invisibles with education e, there is an efficient proportion $p^*(e)$ that should be promoted and an average marginal product $w^*(e)$ that would be associated with the efficiently promoted group. If the quotas and average wages are set to these levels, then a firm can do no better than to promote workers efficiently, since it cannot save on its wage bill by doing otherwise. Then, it is an equilibrium for firms to assign all workers to jobs efficiently, to pay Visibles and promoted workers their marginal products, and to pay unpromoted Invisibles their expected marginal products. To implement this policy, the government must know the functions p^* and w^*, which is a lot to know. It may assume that average productivities are the same among women and minorities as among, say, white males, but if this is wrong (i.e., if women and minorities are in fact either more or less productive than white males), inefficiencies will result. With the existing low and inappropriate stock of human capital among the disadvantaged groups, it is likely that the productivity of older workers in these groups will be lower than for white males; certainly there is no reason to presume equality of skills among the older generation of workers.

Now suppose that the program does set a promotion quota and wage scale designed to raise the average status and pay of women and minorities to the same level as for other groups. Suppose, too, that the labor market moves to a new equilibrium. This equilibrium still provides too little incentive for Invisibles to acquire unobservable human capital, since the wage paid to those in low level jobs does not vary with ability. Full efficiency is not reached. Nevertheless, despite the absence of any scientific criterion to compare the desirability of two equilibria, the new equilibrium appears to us far better in terms of equity, efficiency, and the welfare of Invisibles than the initial equilibrium.

There are some interesting dynamics[11] that result from in applying the combined wage/quota option. When the new affirmative action policy is first introduced, given the pre-existing promotion policy and resulting patterns of human capital acquisition, the average skill level of the female and minority managers is less than that of white male managers. Imposing quotas and common wages is thus likely to meet with much resistance founded on an accurate perception that the beneficiaries of the affirmative action policy are unqualified and overpaid. In the long run, however, if the actual distribution of abilities is the same across all affected groups, the new pay and promotion policies will lead to more similar patterns of human capital acquisition across groups[12], so that the actual productivity of disadvantaged workers will rise to justify the higher wages and more frequent promotions. Thus, in the long run so long as we maintain the policy, wages will end up being in line with skill level. An interesting prediction of the theory is that if the wage-quota policy were abandoned after some years of enforcement and if the underlying Invisibility Problem remained unresolved, the system would return to its old equilibrium in which Invisible workers were inefficiently assigned.

It is possible to construct a private wage contract equilibrium which essentially duplicates the result of the government-imposed wage-and-quota policy described above. In particular, if all firms compete by offering contracts that specify a fixed probability of promotion and an average wage conditional on promotion and if the initial distribution of abilities were the same for Visibles and Invisibles, the resulting equilibrium wages and promotion schedules would be much the same as the one predicted to result from an optimal government-imposed wage-and-quota policy.

Contracts such as we have described do not appear to exist in the private labor market. In part this may be a result of private enforcement costs, since the contract depends on information concerning the wages paid to individuals within the firm, which the firm may be reluctant to reveal. A more fundamental reason is that any individual firm can meet its contractual obligation to promote, say, women or blacks and pay them well by simply hiring the ablest Visible women or blacks at the going wage, without doing anything to improve the lot of Invisibles. Thus, contracts of this form cannot upset the equilibrium described in our theory. Thus, we see no reason to suppose that competition among employers in such an environment would lead to the offering of efficient contracts.

We have thus far observed that it would require an affirmative action program with both a promotion quota and a specified average wage level in order to induce an efficient assignment of the Invisibles workers. Moreover, this two-pronged policy must be imposed continuously in order to be effective. Since this policy is likely to be fairly expensive to administer, we want to consider whether there is a less intrusive policy which would produce an efficient solution.

Rather than affecting firm's behavior directly, consider a class of policies which act directly on the incentives of the individual workers. Part of the efficiency problem in our model is in the level and distribution of human capital investment by the Invisible group. A plausible policy instrument might thus be an education subsidy, available to members of the particular groups who make up most of the Invisible population.

Little can be done to improve the incentives to acquire unobservable human capital. By definition, these acquisition activities are unobserved, so the subsidy cannot be administered in a way that relates directly to the relevant acquisition activities. One can subsidize education, and even offer scholarships for high achievement, but years of education and grades attained are observable to employers as well, so it is best to focus attention on subsidies for observable investments in human capital.

To put the case for educational subsidies in its most favorable light, let us assume that unobservable human capital investments are of negligible importance, that workers do not know their own abilities (so that distortions in the kind of education chosen within the Invisible group are unimportant), that groups can be identified that consist almost solely of Invisibles (so that the subsidies can be well targeted), and that the subsidies are set at a level that leads Invisibles to make the same education choices as Visibles. Then, the distribution of education and skills will be the same for the Visibles and Invisibles, and Proposition 2 applies: Invisibles earn less and are promoted less often than Visibles. Education subsidies may improve the lot of Invisibles, but they can do nothing to eliminate or even ameliorate the basic problem of discrimination by employers.

Finally, we consider whether there are any strategies which individuals in the disadvantaged group could use to reduce the discrimination they experience. One possible answer is for the disadvantaged to try to gain visibility. Indeed this has been a thrust of both the civil rights movements and the womens' movement, as both groups have sought to reduce social discrimination, arguing strongly that such discrimination links up with job and wage discrimination.[13]

A second possibility is in the area of education. It may be that some kinds of education serve directly to enhance visibility. One might then expect members of disadvantaged groups to choose those kinds of education. This may help explain, for example, why minorities receive a disproportionate number of advanced degrees in engineering.

In the well-known Spence labor-market signalling model, education serves as a signal of a worker's ability, and therefore makes the worker visible. Spence's argument uses the "sorting assumption" that, for any smooth schedule of opportunities as a function of education $(\underline{w}(e), \theta^{*}(e), \theta^{**}(e))$, abler workers would choose more education than less able workers. In our model, the sorting assumption may or may not hold, depending on what workers know about their own abilities and on the effect of education on productivity.

Here are two cases in which education fails as a signal. First, suppose workers have no private information about their own abilities; this is completely consistent with our model. Then, clearly, there can be no relation between ability and education. Second, suppose workers do know their own abilities and that education enhances productivity in high level jobs without affecting productivity in low level jobs. Suppose, too, that regardless of the education level attained, a worker the lowest ability level never has a comparative advantage at the high level job. Then, the non-signalling equilibrium specifies a wage schedule for Invisibles assigned to low level jobs that is independent of educational attainment ($\underline{w}(e) = \underline{w}$); only Invisibles of moderate ability acquire education; and employers base promotion decisions for Invisibles on ability only, ignoring education, and end up promoting all those who have acquired education. It is routine to verify that this equilibrium is also a signalling equilibrium. That is, if employers expect the specified behavior from employees and other employers and draw the appropriate inferences, they have no reason to alter their own practices, or to try to hire unpromoted Invisibles away form other employers. Signalling does not seem to be the remedy for the Invisibility Problem.

CONCLUSION

We have shown how non-economic forces arising from segregation and misperceptions can impinge on the functioning of a labor market, resulting in willful discrimination by employers against employees whose backgrounds, personalities, and connections cause their skills to be largely unappreciated by potential employers. It is important to note that, according to our theory, while employers earn rents on the most talented individual Invisibles, they do not earn rents on Invisibles in general. All losses suffered by Invisibles are "dead-weight" losses associated with improper job assignments and inappropriate human capital acquisition decisions: Discrimination results in a loss to society as a whole rather than a transfer from Invisibles to Visibles.

The conclusion that the losses suffered by Invisibles are dead-weight losses has important implications for the politics of implementing affirmative action programs. In the long run, well-designed programs will have a neutral effect on firms, since firms have no economic interest in perpetuating discrimination. In the short run, however, firms may be made worse off by aggressive affirmative action policies, at least until Invisible workers can acquire the skills needed for their new assignments. Thus firms are likely to be most opposed to short time-table programs, but may be neutral toward more gradual approaches. Similarly, since only economy-wide changes can be effective for eliminating discrimination whereas action directed at individual firms can impose losses on them, firms can be expected to resist programs which are narrowly directed.

Among the predictions of our theory are that wages are more variable among Invisibles than among Visibles and that Visibles are more able to increase their wages by investing in human capital. If those predictions are corroborated, and if women, blacks and members of certain ethnic groups are more likely than others to be Invisible, they may help to explain some aspects of the competition for mates and how that competition varies by race. The greater wage dispersion among men than women increases the relative importance of earning capacity as a characteristic of men that attracts women. The man's greater ability to affect his personal earning capacity by his own efforts further contributes to the emphasis among men on finding a well-paying job or career. For women, differences in earning power are less marked, and hence less important for attracting a suitable mate, and women may be less able to raise their incomes by personal effort anyhow. Competition among women for men must then be directed along other dimensions than earning power[14]. Among minorities, where the male wages are predicted to be lower, less variable, and less easily raised by diligence than for more advantaged groups, men will be more likely to compete for women along dimensions other than earning power.

Our explanation of discrimination is not intended to be self-contained. We believe that even job discrimination has non-economic aspects. The puzzle has been why market forces, which are so powerful in many other contexts in ensuring that resources are well used and paid their marginal products, are unable to ensure that women, minorities, and other disadvantaged workers are efficiently employed and paid according to productivity. We have found that market forces are not only weakened by the Invisibility Problem, they are actually twisted into servants of discrimination.

END NOTES

[1] This observation does not apply with full force to education. The years of education acquired, the subjects studied, and the grades attained are all easily observed by potential employers. The actual job skills and work habits acquired, however, are much less easily observed.

[2] By introducing different classes of jobs into the Lundberg-Startz model, one could obtain a theory with nearly the same predictions as ours concerning wage levels and job assignments at the equilibrium of an unregulated labor market. Disadvantaged workers would be inefficiently assigned to jobs because employers are too poorly informed to assign them more efficiently. Such policy tools as quotas, wage restrictions, and educational subsidies are helpless for correcting this problem. In contrast, our theory predicts that properly constructed affirmative action policies will lead to improved job assignments.

[3] This idea -- that an employer gains some ex post monopoly power over its employees by having superior information about them -- has also been explored by Greenwald. His main idea is that an informational disadvantage makes potential employers wary of bidding away another employer's workers, for fear of suffering adverse selection.

[4] The assumption of no cuts and no layoffs can be derived from a more detailed model in which there are large firm specific training costs. Our assumption that contracts simply specify a wage, rather than, say, a wage and a promise that a certain

proportion of women and minorities will be promoted, is more restrictive. Such promises might be made directly, in the employment contract, or indirectly, through a reputation that the firm acquires from its actual behavior toward employees. The consequences of such promises will be explored later in the paper.

[5] In our mathematical model, we treat the level of general education acquired by a worker as perfectly observable, and the benefits of education as perfectly certain.

[6] We limit attention here to the case where the optimal assignment does depend on the worker's ability. Thus, we assume that in the support of θ given e that are some ability levels higher, and some lower, than $\theta^*(e)$.

[7] Of course, at equilibrium, the promotion policy determined in this way together with the market clearing conditions determine x and y, which in turn determine f_1 and f_2. An equilibrium can be found as a fixed point of the composite mapping from R^2 to R^2 taking (f_1,f_2) through θ^*, θ^{**}, $\underline{w}$, p, and q (equations (5)-(9)) and then using the definitions and market clearing conditions (1)-(3) to (x,y) and finally to $(f_1(x,y),f_2(x,y))$. Our assumptions imply that the mapping is continuous, and it is then straightforward to verify that the equilibrium exists.

Later, when the human capital acquisition choices have been made endogenous, $g(e,\theta)$ and $h(e,\theta)$ will be found to depend on w and $\underline{w}$, and the existence argument changes in two ways. First, the basic model must be extended to accommodate the possibility that the joint densities g and h may not exist: the Riemann integrals can be replaced with Riemann-Stieltjes integrals to allow for general distributions. Second, we must allow the possibility of nonconvexities in the acquisition of education. It would be a terrible restriction in our model to assume that the technology for acquiring education is convex; indivisibilities can be a simple consequence of the fact that a worker occupies either a low level job or a high level one. Consequently, we rely on the assumed existence of a continuum of workers to convexify the problem in order to establish the general existence of equilibrium. The techniques required are all standard ones.

[8] By a stochastic upward shift, we mean a shift up in the sense of strict first order stochastic dominance. This means that for any non-constant and nondecreasing function g, $\int g(x)\, dF(x|\varepsilon)$ is strictly increasing in ε.

[9] The expected difference in wages between a Visible and an Invisible investing ε in human capital is

$$\int [w(e,\theta)-\underline{w}(e)]\, dF(\theta|\varepsilon)$$

which is strictly increasing in ε (see note [8]).

[10] This corresponds well with the common wisdom among female managers that it is unwise to acquire good typing skills, because good typists are inevitably assigned to do the typing.

[11] These are market dynamics in the old-fashioned sense of analyzing short-run and long-run equilibrium. Our model is purely static.

[12] There will still be some incentive for Invisibles to underinvest in skills that enhance productivity in the low level jobs, including general skills that enhance productivity in both kinds of jobs.

[13] For example, concerns about visibility are a fundamental part of the public policy issue in the lawsuit against the all-male Jaycees business club.

[14] The old custom of families providing dowries to their marriageable daughters lends credence (if any is needed) to the idea that wealth contributes to the desirability of women.

REFERENCES

Gary S. Becker, The Economics of Discrimination, Chicago: University of Chicago Press, 1957.

Bruce C. Greenwald, "Adverse Selection in the Labor Market," Economics Discussion Paper #166, Bell Laboratories, undated.

Rosabeth M. Kanter, Men and Women of the Corporation, New York: Basic Books, Inc., 1977 (especially page 99).

Shelly J. Lundberg and Richard Startz, "Private Discrimination and Social Intervention in Competitive Labor Markets," American Economic Review, Vol. 73, 1983, 340-347.

Edmund S. Phelps, "The Statistical Theory of Racism and Sexism," American Economic Review, Vol. 62, 1972, 659-661.

A. Michael Spence, "Job Market Signalling," Quarterly Journal of Economics, Vol. 87, 1973, 355-374.

Finis Welch, "Affirmative Action and Its Enforcement," American Economic Review, Papers and Proceedings, Vol. 71, 1981, 127-133.

APPENDIX
Proof of Proposition 1

We break the firm's optimization problem into two parts: (i) wage cost minimization given the levels x and y of high level and low level labor to be used and (ii) profit maximization using the result of the cost minimization problem.

Let $W(x,y)$ denote the minimum total wage cost of acquiring x units of low level labor and y units of high level labor, where the firm's instruments are its hiring policy functions m and n, and its promotion policy functions p and q. Fix the numbers x and y, both strictly positive. We perform a change of variables to work directly with the numbers (rather than proportions) of each kind of worker assigned to each kind of job. Thus, we formulate the mathematical problem as one of minimizing wage costs by choice of the variables $m(e)$ and

$$a(e,\theta) = n(e,\theta)[1-p(e,\theta)]$$

$$b(e,\theta) = n(e,\theta)p(e,\theta)$$

$$c(e,\theta) = m(e)[1-q(e,\theta)]h(\theta|e)$$

The mathematical statement of the problem is:

$$\text{Min} \iint [(a(e,\theta) + b(e,\theta))\, w(e,\theta) + c(e,\theta)\, \underline{w}(e)]\, d\theta\, de$$
$$+ \iint (m(e)h(\theta|e) - c(e,\theta))\, \max[\underline{w}(e), w(e,\theta)]\, d\theta\, de$$

subject to

$$\iint \alpha(e,\theta)\, [a(e,\theta) + c(e,\theta)]\, d\theta\, de \geq x$$

$$\iint \alpha(e,\theta)\, \beta(e,\theta)\, \{b(e,\theta) + m(e)\, h(\theta|e) - c(e,\theta)\}\, d\theta\, de \geq y$$

$$m(e)\, h(\theta|e) - c(e,\theta) \geq 0 \quad \text{for all } (e,\theta)$$

$$a,b,c,m \geq 0$$

We will say that a solution uses the assignment policies of Proposition 1 if $a(e,\theta)$ is zero for $\theta > \theta^*(e)$, $b(e,\theta)$ is zero for $\theta < \theta^*(e)$, $c(e,\theta)$ is zero for $\theta \in (\theta^*(e), \theta^{**}(e))$, and $c(e,\theta)$ is equal to $m(e)\, h(\theta|e)$ otherwise. (It is easy to show that feasible solutions of this form exist -- indeed hiring just Visibles and assigning them according to Proposition 1 leads to a whole class of such solutions.)

Let f_1 (respectively, f_2) denote the minimum cost incurred per unit in hiring a Visible for a low level (resp., high level) job. It is clear by inspection that the firm will be willing to hire positive quantities of each kind of Visible worker only if the Visibles' market wage function takes the form given by (4). Since market clearing requires that the firm hire positive amounts of every type of worker, the equilibrium wage function for Visibles must take the form given by (4).

The requirement that there be positive demand for Visibles of all types also imposes a lower bound on the the equilibrium wage $\underline{w}(e)$ paid to Invisibles occupying low level jobs: It must not be strictly cheaper to hire Invisibles than Visibles. We assume that $\underline{w}(e)$ is chosen so that this requirement is met.

Next we establish that, for any market wage function of the specified form, it is optimal for the firm to use the assignment policies described by (6)-(9), where $\underline{w}(e)$ in (9) is the as yet unspecified market wage function for Invisibles. Indeed, any hiring strategy that uses those assignment policies and that makes the first two constraints hold tightly is cost-minimizing.

As now formulated, the cost-minimization problem has a linear objective and linear inequality constraints; it is a continuous linear program. To prove the optimality of a solution of the specified form, it therefore suffices to identify Lagrange multipliers (shadow prices) λ_x, λ_y, $\mu(e,\theta)$ for the constraints (execept the non-negativity constraints) such that the Lagrange multipliers are dual feasible, and the complementary slackness conditions of linear programming hold. "Dual feasibility" means that the Lagrange multipliers must satisfy constraints (A1)-(A5) below. In what follows, we denote the wage max[$\underline{w}(e), w(e,\theta)$] paid to a high level Invisible by $M(e,\theta)$.

$$w(e,\theta) - \lambda_x\, \alpha(e,\theta) \geq 0 \tag{A1}$$

$$w(e,\theta) - \lambda_y\, \alpha(e,\theta)\, \beta(e,\theta) \geq 0 \tag{A2}$$

$$\underline{w}(e) - M(e,\theta) - \lambda_x\, \alpha(e,\theta) + \lambda_y\, \alpha(e,\theta)\, \beta(e,\theta) + \mu(e,\theta) \geq 0 \tag{A3}$$

$$\int [M(e,\theta) - \lambda_y\, \alpha(e,\theta)\, \beta(e,\theta) - \mu(e,\theta)]\, h(\theta|e)\, d\theta > 0 \tag{A4}$$

$$\lambda_x \geq 0,\ \lambda_y \geq 0,\ \mu(e,\theta) \geq 0 \tag{A5}$$

"Complementary slackness" means that λ_x, λ_y, and $\mu(e,\theta)$ can be positive only if the corresponding constraint in the problem is tight. Also, (A1) must hold with equality unless $a(e,\theta)$ is zero; (A2) must hold with equality unless $b(e,\theta)$ is zero; (A3) must hold with equality unless $c(e,\theta)$ is zero; and (A4) must hold with equality unless $m(e)$ is zero.

The following choice of Lagrange multipliers satisfies these conditions:

$$\lambda_x = f_1 \qquad (A6)$$

$$\lambda_y = f_2 \qquad (A7)$$

$$\mu(e,\theta) = \max(0, \gamma(e,\theta)) \qquad (A8)$$

where

$$\gamma(e,\theta) = M(e,\theta) + f_1\, \alpha(e,\theta) - f_2\, \alpha(e,\theta)\, \beta(e,\theta) - \underline{w}(e) \qquad (A9)$$

It is straightforward to check that $\mu(e,\theta)$ is zero for $\theta \epsilon (\theta^*(e), \theta^{**}(e))$, and positive on the complementary set.

Using (4) and (A6)-(A8), it is direct that (A1)-(A3) hold. Inequality (A4) is precisely the assertion that, on average, Invisibles are paid at least their average marginal product in the jobs they occupy, alluded to earlier. It holds by the assumption that the wages are equilibrium wages and that there is positive demand for each type of Visible.

Given the specified Lagrange multipliers, the complementary slackness (c-s) conditions for the first two constraints of the linear program require that the first two constraints be tight and that $c(e,\theta) = m(e)\, h(\theta|e)$ whenever $\mu(e,\theta)$ is positive, which it is outside the interval $(\theta^*(e), \theta^{**}(e))$. For (A3), c-s requires that $c(e,\theta) = 0$ on the interval. In view of the form of $w(e,\theta)$, c-s for (A1)-(A2) require the assignment policy for Visibles specified in Proposition 1. Taking $m(e) = 0$, c-s is satisfied for (A4). Hence, the solution we have specified is indeed optimal.

Market clearing requires that there exist an alternative optimum in which some Invisibles are hired. That is consistent with complementary slackness for (A4) only if (A4) holds with equality. The unique choice of $\underline{w}(e)$ that allows (A4) to hold with equality is the one given by (5) of Proposition 1.

Thus, we have shown that the specified hiring and assignment policy minimizes wage costs for any fixed (x,y)-target, and that wages of the form given by (4) and (5) are necessary for equilibrium.

The gradient of the wage cost function W(x,y) at any point (x,y) is the vector of shadow prices of the first two constraints. In view of the fact that (A6)-(A7) hold for all positive values of x and y, the wage cost function must be linear and, indeed, must be given by: $W(x,y) = f_1x + f_2y$. Hence the firm's profit maximization problem is:

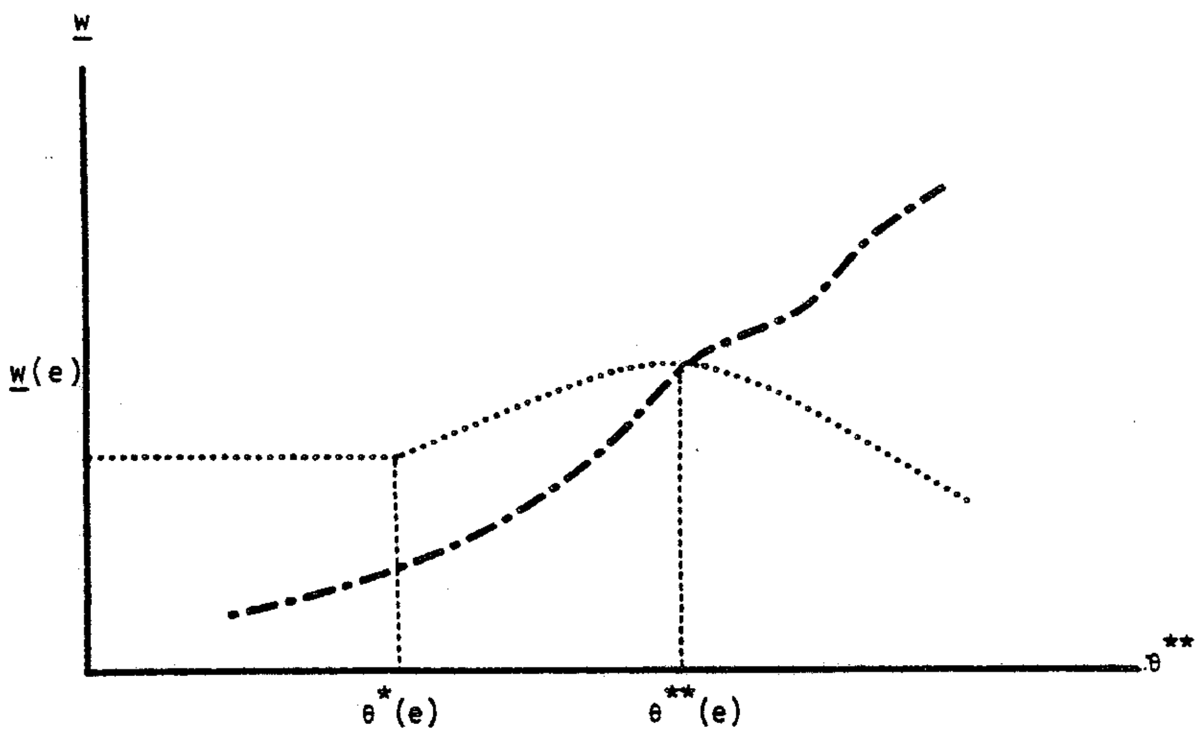
$$\max_{x,y} f(x,y) - f_1 x - f_2 y$$

The optimum has $\partial f/\partial x = f_1$ and $\partial f/\partial y = f_2$.

We have shown that every equilibrium must satisfy (4)-(7) using the parameter values specified. Thus, Propositon 1 is proved.

FIGURE 1

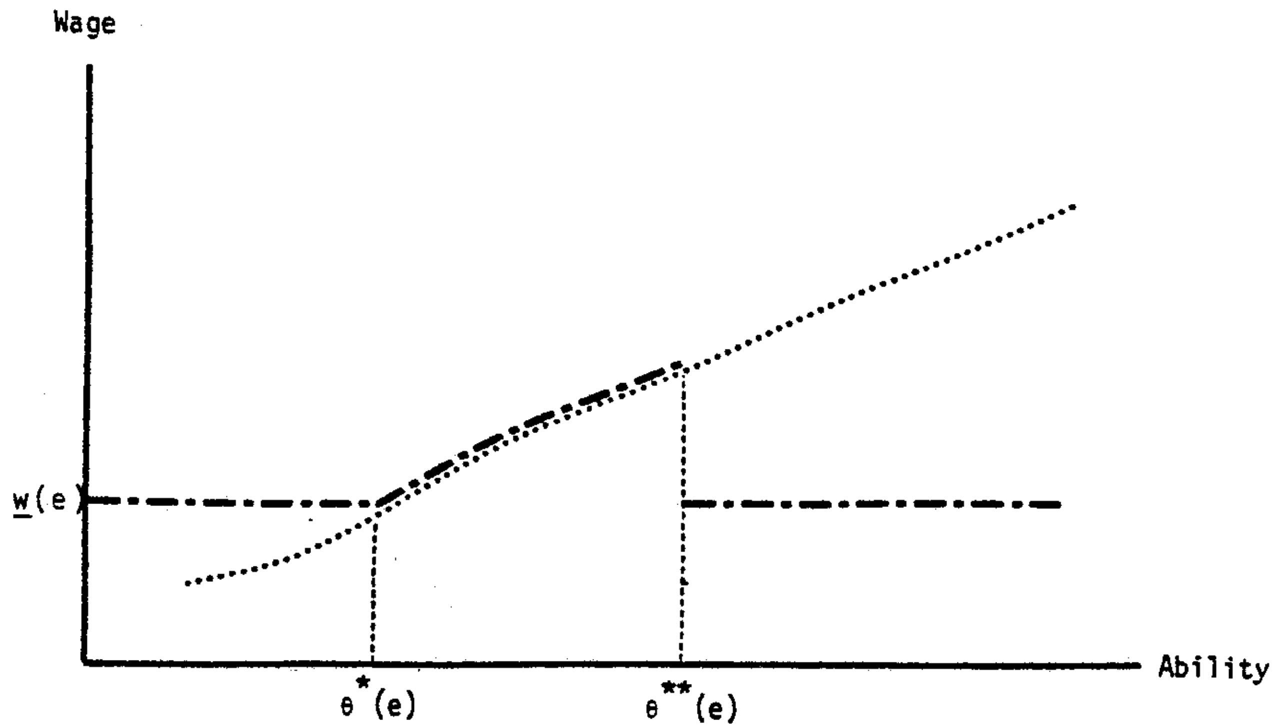
Determination of $\theta^{**}(e)$ and $\underline{w}(e)$



The increasing (dashed) curve shows the relation between θ^{**} and $\underline{w}$ determined by equation (9). The dotted curve shows the relation implied by (5) and (7). The Figure is drawn to show the case in which the second curve is rising at $\theta^{*}(e)$, in which case Invisibles with abilities in the non-empty range from $\theta^{*}(e)$ to $\theta^{**}(e)$ are assigned to high-level jobs. If the second curve is falling at $\theta^{*}(e)$, then the theory predicts that no Invisibles with education e are promoted.

FIGURE 2

Wage as a Function of Ability θ



The dotted line shows the wage earned by a Visible as a function of ability. The dashed line shows the corresponding wage function for Invisibles. The important fact to note here is that the difference between the two wage functions (Visible minus Invisible) is nondecreasing and non-constant.