Eleven

Conceptual Problems in the Enforcement of Anti-Discrimination Laws

GLENN LOURY

THIS CHAPTER considers some problems that arise for anti-discrimination enforcement due to the limited availability of information to an enforcement agent. I intend to stress that under these circumstances enforcement efforts, if not carried out properly, can be counterproductive. Specifically, employment quotas for a group of persons thought to be victims of discrimination can alter employers' and workers' incentives so as to produce undesirable results. Because employers' hiring decisions and workers' investment decisions each depend on perceptions of how the other behaves, one needs an equilibrium model of labor-market interactions to study the issues with which I am concerned. Two such models are provided in what follows.

These models correspond to the two theories of employment discrimination that are most prominent in the literature. The first theory, proposed by Gary Becker (1957), posits that some employers, harboring a "taste" for discrimination against some groups of workers, experience disutility from hiring them. The result can be reduced labor-market opportunities for these workers. The second theory, introduced by Kenneth Arrow (1973), postulates that employers treat some workers differently than others because of disparate statistical generalizations about worker productivity in the various groups. When an individual worker's productivity is not observable, employers may use the information contained in group averages to make their decisions. Arrow showed that this reliance on group averages can lead to discrimination against one group in favor of another, even when the objective capacities of the groups are the same and when the employers hold no invidious motives.

These two theories do not, of course, exhaust the possible explanations of dicrimination that one might offer. But they are a useful context within which to examine the issues that most concern me. These issues have to do with how laws against employment discrimination can be enforced, and what some unintended consequences of enforcement efforts might be. In practice a government determined to prohibit race or sex discrimination in employment faces a difficult problem. If the regulator had unlimited information he

might, by observing the outcome of every employment decision made by a firm, be able to determine if that firm is using exactly the same criteria to select among applicants from different groups. Absent such information, however, enforcement must rest upon means other than exhaustive observation. One response is to compare the numbers of various groups in a particular firm with their population proportions. If there is a significant divergence, the firm might be asked to account for it. Absent a compelling justification, the difference would be presumed to be evidence of bias.

This technique is implicit in one of the key provisions of the U.S. Civil Rights Act of 1991, which holds any employment practice having a "disparate impact" on women or minorities to be unlawful unless the firm can demonstrate that the practice constitutes a "business necessity" (Epstein 1992). (This method is also commonly used in the private sector, by universities in their admissions decisions, and by government agencies in their procurement decisions.) Goals for the employment of minority and women personnel are set for particular enterprises by reference to population percentages of the various groups, sometimes with the comparison population defined so as to approximate the pool of potential employees with skills relevant to the task in question. Enforcement methods such as this have a quota-like quality, in that they specify target numerical outcomes rather than focusing on the particular selection procedures being used.

Objections can be raised against this type of numerical enforcement policy. If the distribution of skills across groups differs in ways not taken into account when the numerical target is set, the representation of women and minorities in the population may overstate their actual employment rates under nondiscriminatory hiring rules. Firms then bear the burden of proving that any difference in group hiring rates is due to a disparity in group skills, which can be quite difficult to do to the satisfaction of a court. Anticipating this difficulty, firms may instead respond to the enforcement regime by adopting hiring rules *in favor* of women and minorities, so as to avoid costly litigation.

This concern about quotas motivated much of the conservative criticism of the Civil Rights Act of 1991. Yet this criticism misses a basic point: If discrimination is in fact taking place, then it reduces the incentives for the workers being discriminated against to acquire skills. When minority workers expect to face biased hiring rules, their returns from acquiring job-relevant skills are lowered. Thus group disparities in skills may simply reflect the presence of employment discrimination. Hence, a regulator may be justified in placing a "burden of proof" on firms whose workforce exhibits substantial underrepresentation of minorities in certain jobs. The ultimate effects of this burden may be to induce firms that would otherwise discriminate to offer equal opportunities to all workers, which might lead to the elimination of any existing skill disparities across groups.

This is the view taken by many liberal advocates of stronger civil rights laws. Yet this view also overlooks an important point, illustrated by the models to follow. Quota-like enforcement policies do not necessarily move workers' incentives in the right direction. If the policy forces firms (even those who would discriminate in the absence of the quota) to set lower standards for minorities, then minority workers may be persuaded that they can get desired jobs without making costly investment in skills. But if fewer members of some group acquire skills, firms will be forced to continue patronizing them in order to achieve parity. Thus skill disparities might persist, or even worsen, under such policies.

Model 1: Taste Discrimination

What follows is a simple taste discrimination model that illustrates these problems. In the model, discrimination reduces the incentives to invest and hence creates skill disparities between groups. These disparities are not necessarily eliminated by statistical enforcement policies because of the adverse incentives such policies can create. *The model suggests that a* gradual *policy, in which representation targets are ratcheted up through time, will be more likely to eliminate both discrimination and skill disparities than a radical policy which, starting from a situation in which skill disparities are significant, demands immediate proportional representation* (see proposition 3).

Consider a labor market consisting of a large number of firms, each of which hires its workforce from a common population. A large number of prospective employees, drawn randomly from this population, approaches each firm seeking employment. These workers belong to one of two identifiable groups, denoted W or B; λ *is the fraction of* Ws in the population, and hence is also the probability that a given applicant for employment is a W. When a firm encounters a worker it must decide whether to "accept" or "reject" the applicant. A firm engages in discrimination if it uses a different rule when deciding whether to accept Bs than it does when deciding about Ws.

Workers are either "qualified" or "unqualified." Firms observe a worker's qualifications before deciding to accept or reject. An accepted worker gains the gross benefit ω, irrespective of her qualifications; a rejected worker's gross payoff is zero. The monetary return to a firm from accepting a worker is $\varkappa_q > 0$ if she is qualified, and $-\varkappa_u < 0$ if she is unqualified. The return to any firm from rejecting any worker is zero. Thus, absent nonmonetary motives, firms would accept all qualified workers and reject all unqualified ones.

Following Becker, I assume firms have a taste for discrimination in that they experience some psychic cost from accepting a B. This cost is greater

the larger the ratio of Bs to Ws among the pool of acceptees. Specifically, let r be the ratio of accepted Bs to Bs for a firm. For some $\gamma > 0$, if a worker is a B then accepting him has the psychic cost to that firm of γr. This means that the net return of accepting a B worker is $x_q - \gamma r$ if he is qualified, and $-x_u - \gamma r$ if he in unqualified. The payoff parameters x_q, x_u, ω, and γ are exogenous. In particular, firms are not allowed to offer lower wages to Bs as compensation for their psychic costs. The focus here is on the implications of enforcement against discriminatory hiring by firms in a world where equal-pay laws are perfectly enforced.

To complete the description of the model I must consider how workers decide whether or not to become qualified. I assume that prior to encountering a firm, a worker can make some costly investment. Once the investment is made, a worker is qualified; otherwise, she is unqualified. The cost to a worker of this investment varies in the population, but is distributed in the same way among Bs and Ws. Let c denote an individual's investment cost, and $G(c)$ the fraction of workers in either groups with cost no greater than c. A worker makes this investment if the expected return from doing so is no less than her cost. I assume that $G(0) = 0$, $G'(c) > 0$, and $G(\omega) < 1$.

The timing of events in this model is as follows: First, each worker decides whether to invest. Then all workers are randomly matched with firms. Finally, each firm observes the group identities and qualifications of its applicants and makes its acceptance decisions. A worker's decision about investment depends on his cost, c, and the group to which he belongs. A firm's decision about whether to accept a given worker depends on that worker's qualifications and the group to which he belongs. An *equilibrium* is a set of decision rules for all workers and firms such that each is a rational (i.e., return-maximizing) response to the others. I shall assume that firms' taste for discrimination is sufficiently great, in the following sense:

$$\textit{Assumption 1: } \gamma > Max\{\frac{\varkappa_q^2}{4\varkappa_u}; \frac{\lambda\varkappa_q}{2(1-\lambda)}\}. \tag{2.1}$$

The purpose of this assumption will become apparent momentarily.

Let us find the equilibrium of this model in the absence of anti-discrimination enforcement. Consider first the behavior of firms. No firm accepts an unqualified B or rejects a qualified W. Doing so would both lower a firm's monetary returns and increase its psychic costs. I claim that under assumption 1 firms reject unqualified Ws and accept qualified Bs as long as the B/W ratio is not too great. To see this, suppose a firm follows some decision rule that results in the acceptance of n_b Bs and n_w Ws. That firm then has a ratio of Bs to Ws among its accepted workers of $r \equiv \frac{n_b}{n_w}$, and so it incurs the psychic cost $\gamma r n_b = \frac{\gamma n_b^2}{n_w}$ due to its discriminatory taste. Hence its marginal psychic cost of accepting another B is $\frac{2\gamma n_b}{n_w} = 2\gamma r$, while its marginal psychic benefit of accepting another W (thereby reducing the B/W ratio) is γr^2. Since the

direct monetary benefit of accepting a qualified B is $\varkappa_q$, the rational firm accepts an additional qualified B if and only if $\varkappa_q \geq 2\gamma r$. That is, a qualified B is accepted as long as the B/W ratio among accepted workers, r, satisfies the inequality: $r \leq r^* \equiv \frac{\varkappa_q}{2\gamma}$. Furthermore, since the monetary cost of accepting an unqualified W is $\varkappa_u$, it pays for a firm to do so only if $\varkappa_u \leq \gamma r^2$. But we know that firms do not permit r to exceed r^*. So accepting an unqualified W does not pay if $\varkappa_u > \gamma r^{*2} = \frac{\varkappa_q^2}{4\gamma}$, which is assured by assumption 1. We conclude that a rational firm rejects unqualified Bs and Ws, accepts qualified Ws, and accepts a qualified B only if $r \leq r^*$.

Let $\pi_b(\pi_w)$ be the fraction of group B (group W) who invest. Since workers are randomly matched with firms and there are many workers per firm, the Law of Large Numbers implies that the shares of qualified Ws and Bs in a firm's applicant pool are approximately $\lambda\pi_w$ and $(1 - \lambda)\pi_b$, respectively. Let $\bar{r} = \frac{1-\lambda}{\lambda}$ be the ratio of Bs to Ws in the population. By the foregoing reasoning if $\bar{r}\left(\frac{\pi_b}{\pi_w}\right) \leq r$ then a firm can expect to accept all of its qualified B applicants, while if $\bar{r}\left(\frac{\pi_b}{\pi_w}\right) > r^*$ a firm will accept some, but not all, qualified Bs. In fact, one sees readily that if firms behave rationally and assumption 1 holds, then a qualified B is accepted with probability $\delta(\pi_b, \pi_w)$, where:

$$\delta(\pi_b, \pi_w) \equiv Min\left\{\frac{\pi_w r^*}{\pi_b \bar{r}}; 1\right\} \tag{2.2}$$

So Bs are discriminated against in equilibrium if workers qualify at rates such that $\left(\frac{\pi_b}{\pi_w}\right) > \left(\frac{r^*}{\bar{r}}\right)$. That is, even though firms dislike accepting Bs, since it is economically advantageous to do so they act on this preference only when Bs are sufficiently numerous among their qualified applicants.

Now we can see that, under assumption 1, discrimination must occur in the equilibrium of this model. Suppose no discrimination was practiced. Then B and W workers with costs $c \leq \omega$ will invest, the fraction of each group becoming qualified will be $\pi_b = \pi_w = G(\omega)$, and thus $\frac{\pi_b}{\pi_w}$ will equal 1. But assumption 1 implies $\frac{r^*}{\bar{r}} \frac{\lambda\varkappa_q}{2\gamma(1-\lambda)} < 1$ which, by the foregoing argument, means that rational firms will desire to discriminate against Bs. This contradicts the initial presumption that they do not.

On the other hand, if qualified Bs face some probability $\delta < 1$ of being accepted, then each B worker expects the average return $\omega\delta$ from investing. So only those Bs with $c \leq \omega\delta$ invest, implying $\pi_b = G(\omega\delta)$. Since Ws are accepted if and only if they invest we have $\pi_w = G(\omega)$. Therefore, in equilibrium each firm will accept all qualified Ws and some but not all qualified Bs, maintaining a B/W ratio just equal to r^*, and implying a probability of acceptance for qualified Bs equal to δ^*, where δ^* solves the equation:

$$\delta \cdot G(\delta\omega) = \left[\frac{r^*}{\bar{r}}\right] \cdot G(\omega) \tag{2.3}$$

It is easily seen that (2.3) has a unique solution δ, with $0 < \delta^* < 1$. Moreover, δ^* is a decreasing function of λ, and it is an increasing function of $\varkappa_q$ and of λ. That is, discrimination against Bs is greater in equilibrium the greater the psychic cost to firms of accepting them, the greater their percentage of the workforce, and the smaller the economic benefit to firms of employing qualified workers.

Let us now consider the effect of anti-discrimination enforcement efforts in this context. I will assume, for the reasons mentioned above, that the government uses a quota-like enforcement strategy, comparing the aggregate number of workers in each group accepted by each firm with some standard, and finding the firm in violation if its employment ratio of Bs to Ws is not sufficiently high. Specifically, suppose that the regulator selects some target ratio $\hat{r} > r^*$ *and* announces that any firm found with a B/W ratio less than $\hat{r}$ will face costly legal proceedings. Let the anticipated costs of these proceedings be so great that no firm wants to risk being found in violation. Then firms will adapt to this regulatory regime by altering their acceptance rules to ensure that $r \geq \hat{r}$. I will assess the effects of this kind of regulation by analyzing how the equilibrium outcome of our simple model changes under this constraint.

The objective of population proportional representation corresponds to a target $\hat{r} = \bar{r}$. However, a regulator might also want to consider more modest objectives. I conduct the analysis under the assumption that the target is set to increase the representation of Bs, but not beyond their relative number in the population, so $\hat{r}$ lies somewhere between r^* and $\bar{r}$: $r^* < \hat{r} \leq \bar{r}$.

To analyze the impact of this constraint, notice first that if the taste for discrimination γ is too large, and the representation target $\hat{r}$ too severe, then the enforcement policy may cause the market to collapse, with no workers being accepted and none investing. To see this, observe that the constraint $r \geq \bar{r}$ will bind for all firms, since a firm in strict compliance (with $r > \hat{r}$) can reject an additional B while remaining in compliance. With $r > \hat{r} > r^*$ *we* know that $2\gamma r$, the marginal benefit of rejecting another B, exceeds $\varkappa_q$, the loss to the firm of rejecting a qualified B. Now, because the constraint is binding, firms will never accept unqualified Ws. So the firm's problem under the regulatory constraint reduces to choosing how many qualified Ws to accept, with Bs then accepted at a rate sufficient to ensure compliance with the anti-discrimination mandate.

Compliance requires there be $\hat{r}$ Bs for each W among the accepted. Therefore, the net benefit from accepting a qualified W, taking account of monetary and psychic returns and of the regulatory constraint, is $\varkappa_q + \hat{r}\varkappa_q - \gamma\hat{r}^2$ if the marginal B accepted is qualified, and is $\varkappa_q - \hat{r}\varkappa_u - \gamma\hat{r}^2$ if the marginal B is unqualified. Thus, if $\varkappa_q + \hat{r}\varkappa_q - \gamma\hat{r}^2 < 0$ firms will reject all their applicants, and the collapse of the market is assured. To avoid this outcome for all $\hat{r} \in (r^*, \bar{r}]$ requires that $\varkappa_q + \hat{r}\varkappa_q - \gamma\hat{r}^2 > 0$, which amounts to the following:

$$\textit{Assumption 2: } \gamma < \left(\frac{\varkappa_q}{1-\lambda}\right) \cdot \left(\frac{\lambda}{1-\lambda}\right) \tag{2.4}$$

Assumptions 1 and 2 together simply state that firms dislike Bs enough to discriminate against them in the absence of regulation, but not so much as to forgo all operations if required to employ qualified Bs at a rate equal to their presence in the population.

Under assumption 2 firms gain by accepting qualified Ws so long as compliance with the regulation can be maintained by accepting qualified Bs. If the ratio of Bs to Ws among a firm's qualified applicants is less than $\hat{r}$, the firm will need to consider whether it pays to accept any unqualified Bs. From the above discussion, this will be worthwhile if:

$$\varkappa_q > \varkappa_u \hat{r} + \gamma \hat{r}^2 \tag{2.5}$$

Hence, there are two cases of interest, depending on whether $\varkappa_q < \varkappa_u \hat{r} + \gamma \hat{r}^2$ [case 1], or $\varkappa_q > \varkappa_u \hat{r} + \gamma \hat{r}^2$ [case 2]. Under case 1, firms accept the maximal number of qualified Ws consistent with being able to remain in compliance by accepting only qualified Bs. Under case 2, firms accept all qualified Ws and as many Bs as necessary to remain in compliance, even if some unqualified Bs must be accepted. For $\hat{r}$ near $\bar{r}$, equation (2.5) is more demanding than assumption 2. Still, it can be shown that values for the parameters $\varkappa_q$, $\varkappa_u$, and γ exist satisfying equation (2.5) for all $\hat{r} \in (r^*, \bar{r}]$, and satisfying assumption 1, if $\lambda > \frac{1}{2}$.

It is now possible to describe how the government's quota-like enforcement policy alters the equilibrium of the model.

Proposition 1: In case 1 there is a unique equilibrium for each statistical enforcement policy $\hat{r} \in (r^*, \bar{r}]$. In this equilibrium all unqualified workers are rejected, all qualified Ws are accepted, and qualified Bs are accepted with probability $\delta(\hat{r})$, $0 < \delta(\hat{r}) \leq 1$. This probability $\delta(\hat{r})$ exceeds the *laissez faire* acceptance rate δ^*, is strictly increasing in $\hat{r}$, and satisfies $\delta(\bar{r}) = 1$. Indeed, $\delta(r)$ solves the equation: $\delta G(\delta\omega) = \left(\frac{r}{\bar{r}}\right) \cdot G(\omega)$.

Proof: If Bs and Ws invest at rates π_b and π_w then the fractions of a firm's applicants who are qualified Bs and Ws are $(1-\lambda)\pi_b$ and $\lambda\pi_w$, respectively. In case 1 firms accept none of the unqualified, and among the qualified all Ws and some Bs if $\frac{\hat{r}}{\bar{r}} < \frac{\pi_b}{\pi_w}$ and all Bs and some Ws if $\frac{\hat{r}}{\bar{r}} > \frac{\pi_b}{\pi_w}$. But when $\hat{r} \in (r^*, \bar{r}]$ an equilibrium with $\frac{\hat{r}}{\bar{r}} > \frac{\pi_b}{\pi_w}$ is impossible, since accepting all Bs and not all Ws from among the qualified means that $\pi_b > \pi_w$, and so $\frac{\hat{r}}{\bar{r}} > 1$, a contradiction. Thus, equilibrium necessarily entails acceptance from among the qualified of all Ws, and some fraction $\delta = \left(\frac{\hat{r}}{\bar{r}}\right) \cdot \left(\frac{\pi_b}{\pi_w}\right)$ of Bs, where $\pi_w = G(\omega)$ and $\pi_b = G(\delta\omega)$. So, the equilibrium acceptance rate for qualified Bs, $\delta(\hat{r})$, solves: $\delta G(\delta\omega) = \left(\frac{\hat{r}}{\bar{r}}\right) \cdot G(\omega)$. This solution is unique, exceeds δ^* for all $\hat{r} \in (r^*, \bar{r}]$, is increasing in $\hat{r}$, and equals 1 when $\hat{r} = \bar{r}$. ■

So in case 1 the use of a quota-like enforcement strategy is sure to produce desirable results. Setting the target $\hat{r}$ equal to the population ratio $\bar{r}$ implies an outcome with no discrimination and no skill disparity between groups. More generally, a stricter target leads to less discrimination by firms and less of a skill disparity between the groups. Matters are not so comforting in case 2.

Proposition 2: In case 2 the equilibrium described in proposition 1 continues to exist. In addition, however, other equilibria may also exist. In these other equilibria all qualified workers are accepted, all unqualified Ws are rejected, but unqualified Bs are accepted with positive probability. A necessary and sufficient condition for the existence of these additional equilibria is that, for some $\delta \in (0,1)$, $\delta \cdot [1 - G(\omega\delta)] > [1 - \left(\frac{\hat{r}}{\bar{r}}\right) \cdot G(\omega)]$.

Proof: In case 2 firms accept unqualified Bs if and only if this is necessary to remain in compliance when accepting all the qualified Ws. In the equilibrium of proposition 1 compliance can be achieved by accepting all qualified Ws and no unqualified Bs, since $\frac{\pi_b}{\pi_w} \geq \frac{\hat{r}}{\bar{r}}$. So this remains an equilibrium in case 2. We seek to identify other (patronizing) equilibria, where $\frac{\pi_b}{\pi_w} < \frac{\hat{r}}{\bar{r}}$ in case 2. In such an equilibrium, to comply while accepting all qualified Ws, firms must accept all qualified Bs and the fraction $\sigma = [\left(\frac{\hat{r}}{\bar{r}}\right)\pi_w - \pi_b]/(1 - \pi_b)$ of unqualified Bs. But then, a B's return from investing is $\delta = 1 - \sigma = [1 - \left(\frac{\hat{r}}{\bar{r}}\right)\pi_w]/(1 - \pi_b)$. Hence the fractions $\pi_b = G(\omega\delta)$ and $\pi_\omega = G(\omega)$ of Bs and Ws would invest. So a patronizing equilibrium exists if $\delta = [1 - \left(\frac{\hat{r}}{\bar{r}}\right)G(\omega)]/[1 - G(\omega\delta)]$ has a solution for some $\delta \in (0,1)$. A necessary and sufficient condition for this to occur is that $\delta \geq [1 - \left(\frac{\hat{r}}{\bar{r}}\right) G(\omega)]/[1 - G(\omega\delta)]$ for some $\delta \in$ in $(0,1)$. Indeed, if this inequality holds strictly then at least two patronizing equilibria exist. ■

In case 2 and under the above-stated condition the use of a quota-like enforcement policy can lead firms to patronize Bs in equilibrium by accepting some even when unqualified, despite the presence of a distaste for accepting Bs. Firms accept unqualified Bs because it is necessary to do so in order to meet the hiring target. But by doing so, firms act to lower the incentive for Bs to invest, thereby inducing a skill disparity disfavoring Bs. Indeed, the skill disparity may actually widen as a result of the regulator's intervention, as compared to the discriminatory equilibrium without intervention. Note that the condition in proposition 2 will be more easily satisfied when $\hat{r}$ is larger. *Equilibria in which Bs are patronized are more likely to exist when the target is more ambitious.*

Intuitively, what happens in a patronizing equilibrium may be seen by considering how firms react to the initial imposition of some hiring target $\hat{r} > r^*$. Prior to the regulation a "surplus" of qualified Bs exists—that is, more Bs invest than find employment ($\delta^* < 1$). So if the target is modest ($\hat{r} \approx r^*$) firms anticipate meeting it by accepting only qualified Bs. But this raises δ, increasing Bs' incentives to invest. So the new equilibrium is as described in proposition 1, with more Bs accepted and the skill gap narrowed.

However, if case 2 applies and if the target is sufficiently ambitious ($\hat{r} \approx \bar{r}$), then firms perceive a "shortage" of qualified Bs relative to the numbers needed to be in compliance while accepting all qualified Ws. They therefore switch from discriminating against qualified Bs to favoring unqualified Bs. This response can actually lower Bs' incentives to invest, leading to a patronizing equilibrium such as that described in proposition 2.

How likely is it that such an equilibrium would arise? This question can be addressed by imagining successive cohorts of workers interacting with firms over time. A dynamic adjustment process is defined for a given hiring target $\hat{r}$ by assuming that firms hire optimally from the workers in each cohort, while investment decisions in cohort $t + 1$ are based on the acceptance rules applied by firms to cohort t. By iterating this process I trace out a long-run response to the enforcement policy. The fraction of Bs investing initially is denoted by π_b^0. Since, in case 2, firms always accept all of the qualified and none of the unqualified Ws, we know that the constant fraction $\pi_\omega = G(\omega)$ of Ws invest in each cohort.

Let π_b^t be Bs' investment rate on cohort t. If $\frac{\pi_b^t}{\pi_w} \geq \frac{\hat{r}}{\bar{r}}$, firms accept a qualified B with probability $\delta^t \equiv \left(\frac{\hat{r}}{\bar{r}}\right) \cdot \frac{\pi_w}{\pi_b^t}$) to comply, so Bs in cohort $t + 1$ invest at rate $\pi_b^{t+1} = G(\omega\delta^t)$. If $\frac{\pi_b^t}{\pi^w} < \frac{\hat{r}}{\bar{r}}$, all qualified Bs and the fraction $\sigma^t = [(\frac{\hat{r}}{\bar{r}})$ $\pi_\omega - \pi_b^t]/(1 - \pi_b^t)$ of unqualified Bs are accepted, so Bs' in cohort $t + 1$ invest at rate $\pi_b^{t+1} = G(\omega(1 - \sigma^t))$. Thus $\{\pi_b^t\}$, $t \geq 0$, solves:

$$\pi_b^{t+1} = G\left(\omega \cdot \left[\frac{\hat{r}}{\bar{r}}\right] \cdot \left[\frac{\pi_\omega}{\pi_b^t}\right]\right), \qquad \text{for } \frac{\pi_b^t}{\pi_\omega} \geq \frac{\hat{r}}{\bar{r}} \tag{2.6}$$

$$\pi_b^{t+1} = G\left(\omega \cdot \left[1 - \left(\frac{\hat{r}}{\bar{r}}\right)\pi_\omega\right]/[1 - \pi_b^t]\right), \quad \text{for } \frac{\pi_b^t}{\pi_\omega} < \frac{\hat{r}}{\bar{r}} \tag{2.7}$$

The stationary points of this difference equation correspond exactly to the fraction of Bs who invest in the equilibria of our model. Hence, if a patronizing equilibrium exists then equation (2.6) has a stationary point π_b^* at which $\frac{\pi_b}{\pi^w} < \frac{\hat{r}}{\bar{r}}$. Detailed study of equation (2.6) leads to the following result:

Proposition 3: If a patronizing equilibrium exists from enforcement policy $\hat{r} \in (r_\cdot, \bar{r}]$ then one also exists from any $r' \in (r_\cdot, \bar{r}]$. Let $r \equiv \inf\{r \in (r_\cdot, \bar{r}] |$ a pat. equil. exists from r$\}$. If $\frac{\omega G'(\omega)}{1 - G(\omega)} > 1$ then $\underline{r} < \bar{r}$, and $\exists$ a continuous, increasing function $\tilde{\pi}_b : (\underline{r}, \bar{r}] \to (0, \pi_\omega]$ with $\tilde{\pi}_p(\bar{r}) = \pi_\omega$ *such that* $\forall\, \hat{r} \in (\underline{r}, \bar{r}]$, equation (2.6) converges to a patronizing equilibrium whenever $\pi_b^0 < \tilde{\pi}_b(\hat{r})$.

Sketch of Proof: Figure 11.1 graphs the difference equation, equation (2.6). Notice that if a patronizing equilibrium exists at all, then this equation has at least two stationary points that are less than $\frac{\hat{r}}{\bar{r}} \cdot \pi_\omega$. Denote by $\tilde{\pi}(\hat{r})$ the largest of these stationary points. With this definition it is a straightforward exercise to verify the claims in the proposition. ■

To see the implications of this result, note that $\pi b^0 < \pi_\omega$ *necessarily*, or else there is no need for regulation. Now, suppose that $\varkappa_q > \varkappa_u \bar{r} + \gamma \bar{r}^2$ and

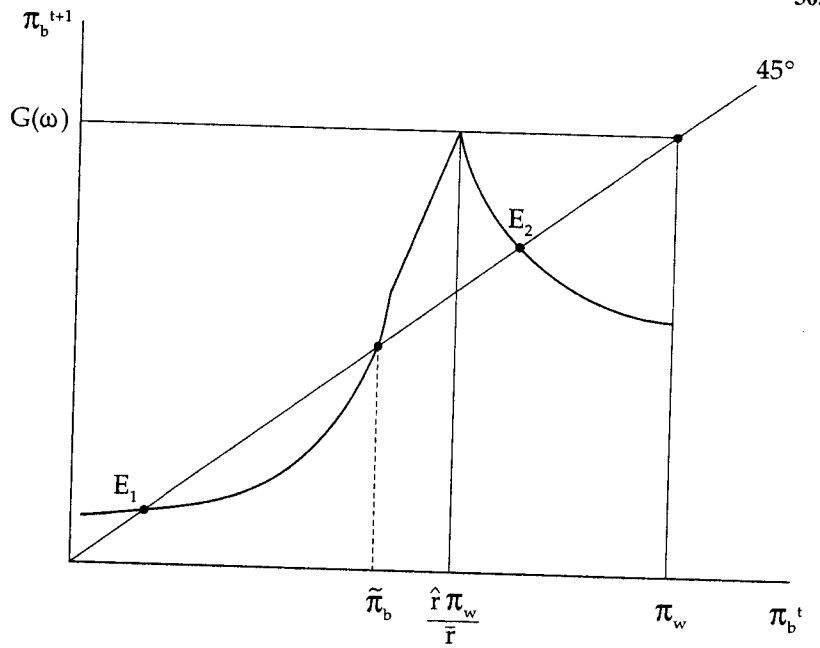

Figure 11.1. E_1 is a (locally stable) patronizing equilibrium.

that $\frac{\omega G'(\omega)}{1 - G(\omega)} > 1$. Then proposition 3 states that *a patronizing outcome is the inevitable result of the regulator seeking fully proportional representation* [$\hat{r} = \bar{r}$]. But since the threshold investment rate $\tilde{\pi}_b(\hat{r})$ is increasing in $\hat{r}$, the regulator could avoid an equilibrium in which Bs are patronized by setting a more modest target. For example, it is easy to see that a target equal to the *laissez-faire* ratio of qualified Bs to Ws [$\hat{r} = \bar{r} \cdot \frac{G(\delta^*\omega)}{G(\omega)}$] never leads to patronization.

These results show that anti-discrimination enforcement may embody an awkward trade-off: A policy of proportional representation of Bs risks inducing a patronizing outcome, whereas a more modest target, though less prone to that problem, will not fully eliminate discrimination. Yet the model also suggests a way around this trade-off. Rather than settling immediately upon the proportional target $\hat{r} = \bar{r}$, the regulator could instead operate a gradual policy, with the target being ratcheted up in a series of steps.

Suppose, for example, that the adjustment process described in equation (2.6) operates quickly relative to the rate at which the policy target is changed. Then, if the regulator always sets the new target r' so that the currently prevailing investment rate among Bs, π_b, satisfies $\pi b > \tilde{\pi}_b(r')$, a patronizing outcome can be avoided. But, by proposition 1, each time the target is raised the rate of investment among Bs improves, which permits the

target to be raised further without risk of patronization. By proceeding in this way, this process can eventually reduce both discrimination and also skill disparities between the groups to an arbitrarily small level (though it can never completely eliminate them).

Model 2: Statistical Discrimination

I now consider a model based on Arrow's theory of statistical discrimination. In this model firms have no desire to avoid hiring *B*s per se. However, because firms cannot observe a worker's qualifications directly, they use group-based stereotypes about workers when making hiring decisions. These stereotypes can become a self-fulfilling prophecy, as explained more fully below. A regulator may then want to intervene so as to break the firms of the habit of relying on stereotypes. *But the model to follow shows that such intervention, when it takes a quota-like form, can be self-defeating* (see proposition 5).

My concern is with the standards employees use to assign workers to desirable positions, the effort workers invest to acquire skills useful in those positions, and the ways in which assignment and investment decisions are affected by group hiring standards. Consider the interaction between an employer and a large number of workers divided into two racial groups, *B*s and *W*s; again λ denotes the fraction of *W*s in this population. The employer observes a worker's color, and so can treat *B* and *W* workers differently. The sole action of the employer is to assign each worker to one of two positions, called task "zero" and task "one." Assignment to task one may be thought of as giving the worker a promotion; it is the more desirable, but also more demanding, of the two positions.

I assume that all workers perform satisfactorily at task zero. Workers decide, before the employer assigns them to a task and without the employer's knowledge, whether to invest in the acquisition of a skill essential for effective performance at task one. The investment is costly for a worker to make. The size of this cost varies from worker to worker, though in a manner that is statistically the same for each racial group. Imagine, for example, that more able workers find it easier to acquire the skill needed for task one, and that the distribution of ability is the same within each group. Let c denote a worker's cost. I assume that each worker's cost is a random draw from a uniform distribution on the interval [0,1], regardless of her group.

The employer cannot observe a particular worker's cost (ability). What he can observe is the group identity of each worker and the outcome of a skills test, to be described momentarily. Although the two groups have the same distribution of ability, they need not exhibit the same pattern of investment. Workers with the same investment cost but belonging to different groups

might make different investment decisions, if they anticipate they will receive different treatment by the employer.

I assume that a worker obtains a premium whenever she gains assignment to task one, whether she has acquired the needed skill or not. However, since an unskilled worker performs inadequately, the employer wants a worker in task one if and only if she has the acquired requisite skill. Otherwise he wants that worker to go to task zero. The employer's profits are greatest when all skilled workers are assigned to task one and all unskilled workers to task zero. Specifically, let ω be the premium a worker puts on assignment to task one. Let $\varkappa_1 > 0$ be the employer's net gain from assigning an investing worker to task one instead of task zero; and let $\varkappa_0 > 0$ be his net gain from assigning a noninvesting worker to task zero rather than task one. Define $r \equiv \frac{\varkappa_1}{\varkappa_0}$.

The employer wants to match workers to their most productive tasks. Lacking any prior information, the employer "tests" a worker's qualification for doing task one. That is, he gathers what information he can (from an interview, analysis of previous work history, written exam, etc.) in order to assess the worker's capabilities. Let θ denote the test outcome. I assume that three test outcomes are possible: (1) it is clear that the worker can do task one (θ = "pass"); (2) it is clear that she cannot (θ = "fail"); and (3) it is uncertain whether the worker can do task one (θ = "unclear"). I assume that if a worker invests she cannot fail the test, and if a worker does not invest she cannot pass the test. Let $p_1(p_0)$ be the probability that an investing (noninvesting) worker gets an unclear test result. So $1 - p_1$ is the probability that an investing worker passes the test, and $1 - p_0$ is the probability that a noninvesting worker fails it. I require the following assumption:

$$\textit{Assumption 3: } p_1 > p_0 > 1 - \frac{1}{\omega} \tag{3.1}$$

Assumption 3 implies that the test is better at revealing noninvestors than investors, but not so good as to induce all workers to invest.

I now consider the behavior of the workers and employer in this model. Let $A(i, \theta)$ denote the probability that the employer assigns to task one a worker from group i with test outcome θ, for $i \in \{B,W\}$ and $\theta \in T \equiv \{pass, fail, unclear\}$. A strategy for the employer is any function $A : \{B, W\} \times T \to [0, 1]$. Let $I(i, c)$ denote the probability that a group i worker with cost c invest in the skill needed to do task one, for $i \in \{B, W\}$ and $c \in [0, 1]$. A strategy for workers is any function $I : \{B, W\} \times [0, 1] \to [0, 1]$. The sequence of events is that each worker, knowing his color and his investment cost, decides whether to acquire the skill needed for task one. The employer then encounters the workers, gives them the test, and, on the basis of the test outcome and (possibly) the worker's color, assigns the workers to a task.

An equilibrium, then, is a pair of functions $< A^*, I^* >$ such that each

strategy maximizes that decision maker's anticipated net reward, given the available information and the strategies employed by the other agents. I show that despite the absence of any racially invidious motive on the part of the employer, discrimination against *B*s can arise in an equilibrium of this model.

To find the equilibria I begin by considering the employer's decision in each contingency. Clearly she assigns anyone passing the test to task one, and anyone failing it to task zero, regardless of color: $A^*(i, pass) = 1$ and $A^*(i, fail) = 0$ in any equilibrium. Let $\alpha_i \equiv A^*(i, unclear)$ denote the equilibrium probability a worker in group i who gets an unclear test outcome is assigned to task one. When $\theta =$ unclear the employer must assess the conditional probability that the worker has invested, $\xi_i \equiv \Pr[I^*(i, c) = 1 | \theta = unclear]$, in order to decide which assignment is best. Her net expected return from putting the worker in task one rather than task zero is $\xi_i \varkappa_1 - (1 - \xi_i)\varkappa_0$. Thus, if $\xi_i > (\frac{\varkappa_0}{\varkappa_0 + \varkappa_1})$, then we must have $\alpha_i = 1$, while if $\xi_i < (\frac{\varkappa_0}{\varkappa_0 + \varkappa_1})$ then we must have $\alpha_i = 0$. Note that $\frac{\varkappa_0}{\varkappa_0 + \varkappa_1} = \frac{1}{1+r}$.

Given an unclear test result, the odds that the worker producing it has invested depend on the mean rate of investing in that worker's group and the likelihood that investors and noninvestors get unclear results. Let $\pi_i \equiv \int_0^1 I^*(i, c)dc$, the mean equilibrium investment rate in group i, $i \in \{B, W\}$. Bayes's Rule then implies that $\xi_i = \frac{\pi_i p_1}{\pi_i p_1 + (1 - \pi_i)p_0}$. We conclude that there exists a threshold investment rate $\pi^* such$ that, in equilibrium, $\pi_i > \pi^*$ implies $\alpha_i = 1$, and $\pi_i < \pi^*$ implies $\alpha_i = 0$, where $\pi^* \equiv \frac{p_0}{p_0 + p_1 r}$. We may interpret this result as follows: For a given group of workers, only if the employer believes the fraction of investors is sufficiently large will he give any one of them the "benefit of the doubt" in the face of an unclear test outcome. Define the employer's best response correspondence, $\phi_e(\pi)$, as follows:

$$\alpha_i \in \phi_e(\pi_i) \equiv \begin{cases} \{1\} & \text{if } \pi_i > \pi^*, \\ [0, 1] & \text{if } \pi_i = \pi^*, \\ \{0\} & \text{if } \pi_i < \pi^* \end{cases} \tag{3.2}$$

I call the employer "optimistic" about group i when $\pi_i \geq \pi^*$, and "pessimistic" when $\pi_i < \pi^*$. I call the employer "liberal" toward group i if $\alpha_i = 1$, and "conservative" if $\alpha_i = 0$. I say the employer "discriminates" against *B*s in a given equilibrium if he is conservative toward *B*s while being liberal toward *W*s.

To see how equilibria with discrimination can occur in this model, we must consider the workers' behavior. A worker invests only if she expects the gain to exceed the cost. Since in any equilibrium passing workers always gain assignment to task one and failing workers never do, the key issue for workers contemplating whether or not to invest is what happens if the test outcome is unclear. It is easy to see that $I^*(i, c) = 1$ in equilibrium if $c < c_i^*$,

and $I^*(i, c) = 0$ if $c > c_i^*$, for $c_i^* \equiv \omega \cdot [\alpha_i(1 - p_0) + (1 - \alpha_i)(1 - p_1)]$. So $\pi_i = \int_0^1 I^*(i, c)dc = c_i^*$ in any equilibrium. Define the workers' best response function, $\phi_\varphi(\alpha)$, as follows:

$$\pi_i = \phi_\omega(\alpha_i) \equiv \omega \cdot [\alpha_i(1 - p_0) + (1 - \alpha_i)(1 - p_1)] \tag{3.3}$$

If a group of workers expect liberal treatment from the employer then $\alpha_i = 1$, and the fraction $\pi_l \equiv \omega(1 - p_0)$ invest. If a group of workers anticipates conservative treatment then $\alpha_i = 0$, *and the fraction* $\pi_c \equiv \omega(1 - p_1)$ invest. Assumption 3 implies that $0 < \pi_c < \pi_l < 1$. Our earlier analysis shows that in any equilibrium the following relationship obtains:

$$\alpha_i \in \phi_e(\phi_\omega(\alpha_i)),\ i \in \{B, W\} \tag{3.4}$$

This equation states that the employer's behavior is optimal toward each group of workers, given their respective mean investment rates, and at the same time workers' investment decisions are optimal, given the employer's behavior toward their group when the test outcome is unclear. We adopt the following assumption:

$$\textit{Assumption 4: } \frac{1}{\omega(1 - p_0)} < 1 + \frac{rp_1}{p_0} < \frac{1}{\omega(1 - p_1)}$$

Proposition 4: Given assumption 3, a discriminatory equilibrium exists in which the employer is optimistic about and liberal toward *W*s, who invest at rate π_l, while being pessimistic about and conservative toward *B*s, who invest at rate π_c, if (and, with weak inequalities, only if) assumption 4 holds.

Sketch of Proof: Figure 11.2 graphs the best response relations $\phi_e(\pi)$ and $\phi_\omega(\alpha)$ in the (α, π) unit square. Then, in view of equation (3.4), the result should be obvious. ■

Assumption 4 is simply the requirement that $\pi_c < \pi^* < \pi_l$. This condition requires that r, the relative employer benefit of correctly assigning an investing worker as compared to correctly assigning a noninvesting worker, is neither too large nor too small. In a discriminatory equilibrium the employer, by treating the groups differently in the event of an unclear test outcome, creates unequal incentives for workers in the two groups to become skilled. Although the employer's differential treatment is justified by his (correct in equilibrium) belief that workers in the two groups have unequal mean investment rates, this investment disparity is itself due to his differential treatment. That is, in a discriminatory equilibrium the belief that *B*s are on average less skillful that *W*s is a self-fulfilling prophecy.

In a discriminatory equilibrium, when the employer is not acting in a "color-blind" fashion, it is natural for an anti-discrimination enforcement agent to try to correct this discrimination by forcing the employer to assign workers from each group to each task at the same rate. This enforcement official might proceed in one of two ways. Ideally, she would insist on color-

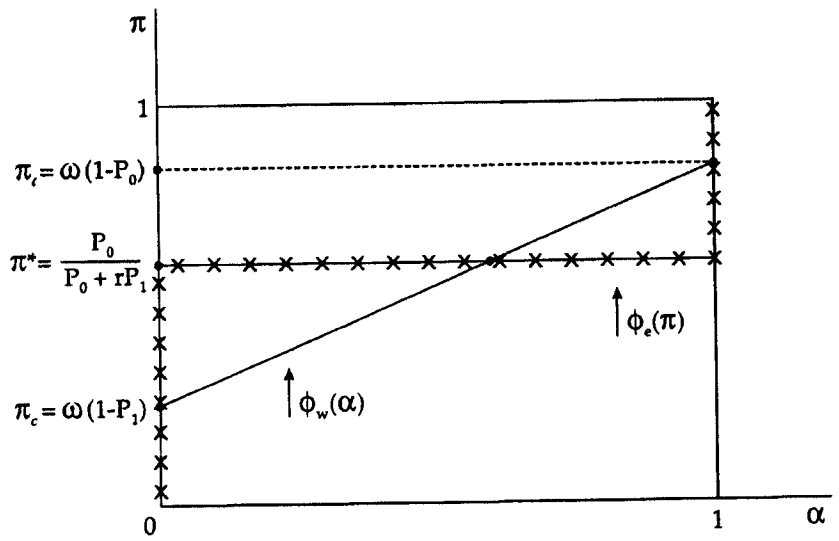

Figure 11.2. Proof that a discriminatory equilibrium exists under A4

blind behavior by forbidding the employer to treat *W*s and *B*s with unclear tests any differently. That is, the official would require $A(B, \theta) = A(W, \theta)$, $\forall \theta \in T$. This would be difficult to implement, however. Consider the informational demands of such a policy. The enforcement agent would have to observe all information upon which an employer might base his assignment (interviews, work history, etc.) to determine if *B*s and *W*s are really being treated the same. In most situations this is not possible.

As a second way of proceeding, the enforcement agent might take a more "results-oriented" rather than "process-oriented" approach. Here, the government monitors the rate at which workers in the respective groups are assigned to various positions, insisting on proportional representation. This quota-like policy, which I will refer to as "affirmative action," leads both the employer and the government to depart from purely color-blind practice. The government must monitor the racial composition of the employer's workforce in each task, while the employer, if she is to comply with the policy, must calibrate her hiring policy, given the workers' investment strategies, so as to anticipate achieving equal proportionate representation. I will now examine in the context of the model set out above whether this intervention eliminates the *B/W* difference in investment incentives that prevailed in the discriminatory equilibrium.

Suppose that we start in a discriminatory equilibrium in which $\pi_\omega = \pi_l$ and $\pi_b = \pi_c$. Let the anti-discrimination enforcement authorities enact a policy requiring that each group be assigned to each task at the same overall

rate. Initially the employer is violating this policy. Indeed, if ρ_i is the rate group i workers are assigned to task one, then

$$\rho_i = \int_0^1 \{\Sigma_\theta \Pr[\theta] I^*(i,c)] \cdot A^*(i,\theta)\} dc, \tag{3.5}$$

so $\rho_\omega = \pi_1 + (1 - \pi_1)p_0 > \pi_l > \pi_c > \rho_b = \pi_c(1 - p_1)$ in the initial situation.

Therefore, in order to comply with the affirmative action mandate the employer must either assign more *B*s or fewer *W*s to task one. Since he is maximizing his profits in the initial equilibrium, both alternatives lower his net payoff. Which course is least undesirable to him, however, depends on the relative numbers of *B* and *B* workers in the population. In general the employer will try to minimize the number of instances where, in order to comply with the enforcement policy, he has to assign a worker of either group to a task that he believes will not be most profitable for him. If *B*s are comparatively few, then by reassigning some *B*s to task one instead of task zero he could meet the enforcement mandate with a relatively small number of unprofitable assignments. On the other hand, if *B*s are numerous in comparison to *W*s, then by reassigning some *W*s to task zero instead of task one, he could meet the government's hiring requirement at least cost to himself.

Specifically, there exists a number $\lambda^* \in (0,1)$ such that when $\lambda > \lambda^*$ the employer always responds to anti-discrimination enforcement efforts by reassigning *B*s from task zero to task one, and never by reassigning *W*s from task one to task zero. To see this, consider reassigning either ΔB more blacks in task one, or alternatively ΔW more whites to task zero, where the object in each case is to reduce the difference in the proportions of black and white workers going into task one by the same amount. So $\frac{\Delta B}{1-\lambda} = \frac{\Delta W}{\lambda}$, or equivalently, $\Delta W = (\frac{\lambda}{1-\lambda})\,\Delta B$. In the initial equilibrium the mean investment rate among *W*s is $\pi_\omega = \pi_l$, so the employer expects to lose $\xi_1 \varkappa_1 - (1 - \xi_1)\varkappa_0$ on each *W* unclear test outcome who is reassigned, where $\xi_1 = \pi_l p_1 / [\pi_l p_1 + (1 - \pi_l)p_0]$. On the other hand, if a *B* who *fails* the test is reassigned to task one from task zero the employer is sure to lose x_0. Hence it is less costly to the employer to accomplish a given increase in the relative representation of *B*s in task one by reassigning *B* and not *W* workers if $\frac{\lambda}{1-\lambda} \cdot [\xi_1 \varkappa_1 - (1 - \xi_1)\varkappa_0]$, which obtains for λ sufficiently large, as long as $\pi_l > \pi^*$. This is formalized below:

$$\textit{Assumption 5: } \lambda > \lambda^* \equiv \left[\, 1 + \left(\frac{p_0}{p_1}\right) \cdot \left(\frac{1 - \pi_l}{\pi_l}\right) \right] / [1 + r] \tag{3.6}$$

An equilibrium under affirmative action is a pair of strategies $\langle I'(i, c), A'(i, \theta)\rangle$ such that workers are responding optimally to A', while the em-

ployer is responding optimally to I' subject to the constraint $\rho_\omega = \rho_b$. This constraint may be stated formally as:

$$\int_0^1 \{\Sigma_\theta \Pr[\theta|I'(B,c)] \cdot A'(B,\theta)\}dc =$$

$$\int_0^1 \{\Sigma_\theta \Pr[\theta|I'(W,c)] \cdot A'(W,\theta)\}dc \qquad (3.7)$$

Starting from the initial discriminatory equilibrium $\langle I^*, A^*\rangle$, the employer would not expect to meet the enforcement agent's equal representation mandate by simply following a color-blind policy, since the initial mean investment rate is smaller among *B*s than *W*s. Moreover, under assumption 5, since the employer does not reassign any *W* workers, her best response, consistent with the anti-discrimination requirements, to the workers' strategy I^* must involve putting *B* workers who fail the test, and who the employer therefore knows are unqualified, into task one. When she does this I say that she "patronizes" *B* workers. Let β denote the probability that a *B* worker who fails the test is nevertheless assigned to task one. I call β the employer's "degree of patronization."

The degree of patronization necessary for the employer to assure equal group representation in task one depends on the mean rates of investment in the two groups. The investment rate among *W*s is fixed at π_l, since no *W*s are going to be reassigned. Let $\pi_b \leq \pi_l$ be given. Equal task one representation requires $\pi_b + (1 - \pi_b)[p_0 + \beta(1 - p_0)] = \pi_l + (1 - \pi_l)p_0$, or:

$$\beta = \frac{\pi_l - \pi_b}{1 - \pi_b} \qquad (3.8)$$

On the other hand, if the employer is liberal toward *B*s with an unclear test outcome, and uses degree of patronization $\beta > 0$ with *B*s failing the test, then the resulting incentive for *B*s to invest will be less than the incentive for *W*s to invest. Any positive degree of patronization lowers a *B* worker's expected gain from investing, compared to merely being treated liberally but not patronized, because a positive degree of patronization raises the chance for a noninvestor to get into task one without affecting the fact that an investor is certain to gain that assignment. Specifically, an investing *B* worker gains ω with probability one; a non-investing *B* gains ω with probability $p_0 + \beta(1 - p_0)$. So investment for a *B* worker is optimal only if $c \leq \omega(1 - p_0)(1 - \beta) = \pi_l(1 - \beta)$. We conclude that when *B* workers are making a best response to the employer's enforcement-influenced assignment strategy, they are investing at a mean rate π_b given below:

$$\pi_b = \pi_l(1 - \beta) \qquad (3.9)$$

Combining equations (3.8) and (3.9), we see that *B*s' mean investment rate in any equilibrium under affirmative action satisfies:

$$\pi_b' = \frac{\pi_l(1 - \pi_l)}{1 - \pi_b'}, \quad \text{and } \pi_b' \leq \pi_l \qquad (3.10)$$

Thus two equilibrium situations are possible under affirmative action: $\pi_b' = \pi_l$, and $\pi_b' = 1 - \pi_l$. The second situation occurs if and only if $\pi_l > \frac{1}{2}$. The first situation has an obvious interpretation. Should the employer come to believe that *B*s are investing at rate π_l, the same as *W*s, he would want to be liberal but not patronizing toward them, and would comply with the government's mandate by doing so. If *B*s expect the degree of patronization to be zero, they, like *W*s, would invest at rate π_l. When this equilibrium arises the employer's initial discriminatory beliefs have been eliminated by the use of the affirmative action enforcement tool. The government's insistence on equal representation for each group creates a situation in which the opportunities, and so the distributions of skills, for each group of workers are equalized. Having achieved this result, affirmative action policy can "wither away" because the employer's discriminatory beliefs, which justified (for him) the initial unequal treatment of *B*s, have been dispelled.

A second situation has a rather less obvious, but no less important, interpretation: the employer continues to think *B*s invest at a lower mean rate than *W*s. She therefore persists in patronizing them to some degree. But because *B*s when patronized have a lower incentive to invest than *W*s, the employer's belief that patronization is needed becomes a self-fulfilling prophecy. In this instance, rather than creating equality of opportunity, the enforcement policy leads to a situation in which, in order to meet the equal representation requirement, the employer discriminates in favor of unskilled *B*s. Because noninvesting *B*s have superior opportunities, the return to acquiring a skill is lower for *B*s than *W*s, and relatively fewer *B*s invest. So the employer has to continually favor *B* workers in order to comply with the government's mandate. In this equilibrium affirmative action, far from "withering away," sets in motion a sequence of events that guarantee that it may have to maintain indefinitely.[1] The incentives for the employer, and hence for *B* and *W* workers, are altered by the government's use of a color-conscious enforcement strategy in such a way that a group difference in workers' acquisition of skills is sustained.[2] This is precisely the unintended negative consequence of racial preferences to which I alluded in the introduction.

[1] This conclusion is true only if the mean investment rate among *B*s in this second equilibrium is low enough that the employer would want to be conservative toward them were he not constrained to meet the government's mandate, that is, if $1 - \pi_l < \pi^*$.

[2] The *B*s' skill acquisition rate in this equilibrium ($\pi_b' = 1 - \pi_l$) could even turn out to be smaller than in the initial discriminatory equilibrium (π_c). Thus $1 - \pi_l < \pi_c$ *if* $\pi_l + \pi_c > 1$, or equivalently, if $\omega(2 - p_0 - p_1) > 1$. Hence if the worker's value of getting task one is big enough, and/or the test is sufficiently accurate, then this extreme illustration of the "law of unintended consequences" will, in fact, obtain in this model.

It is therefore of some interest to determine which of these two equilibria under affirmative action will actually obtain. At the initial discriminatory equilibrium the employer thinks she needs some patronization, but her use of it alters *B*s' investment incentives. As *B* workers change their behavior, the degree of patronization the employer believes to be necessary also changes. Imagine an adjustment process, similar to that employed in the study of the previous model in this chapter, in which the employer and *B* workers alter their behavior over a sequence of stages, each party reacting to the behavior observed from the other at the previous stage of adjustment. It is plausible to postulate that the equilibrium reached under affirmative action is the one that eventually emerges from this iterative process of adjustment.

Proposition 5: If $\pi_l > \frac{1}{2}$, and if assumptions 3–5 hold, then the adjustment process described above converges monotonically to an equilibrium under affirmative action in which the degree of patronization is positive.

Proof: Let π_b^t be the mean investment rate among *B*s at stage t of this adjustment process, $t = 1,2, \ldots$ At stage one $\pi_b^1 = \pi_c < \pi_l$. To comply with the affirmative action mandate the employer adopts that degree of patronization β^t which, given π_b^t, satisfies equation (3.8): $\beta^t = \frac{\pi_l - \pi_b^t}{1 - \pi_b^t}$. Given degree of patronization β^t, the mean investment rate of *B*s at stage $t + 1$, using equation (3.9), is $\pi_b^{t+1} = \pi_l(1 - \beta^t)$. Combining these results leads to the difference equation:

$$\beta^{t+1} = \frac{\beta^t}{\left(\frac{1-\pi_l}{\pi_l}\right) + \beta^t}, \quad \beta^1 = \frac{\pi_l - \pi_c}{1 - \pi_c} \tag{3.11}$$

It is straightforward to verify the following: If $\pi_l \leq \frac{1}{2}$, then $\{\beta^t\} \to 0$ and $\{\pi_b^t\} \to \pi_l$ as $t \to \infty$. If $\pi_l > \frac{1}{2}$, then $\{\beta^t\} \to 1 - \left(\frac{1-\pi_l}{\pi_l}\right) > 0$, and $\{\pi_b^t\} \to 1 - \pi_l$ as $t \to \infty$. ■

Another way of saying this is that *the undesirable outcome obtains under affirmative action if, when facing a liberal employer, the average worker would strictly prefer to invest in the skill needed for task one.* The average worker will want to invest when facing a liberal employer only if the expected return from doing so exceeds his investment cost. This expected return is greater the greater the gain to a worker from being assigned to task one, and the lower the probability that a noninvestor goes undetected by the test. Thus, the higher the value of assignment to task one, relative to the average worker's investment cost, and the more powerful the test for identifying noninvestors, the more likely that a patronizing equilibrium will arise under affirmative action. The patronizing outcome is also more likely when the disadvantaged group is a relatively small fraction of the total population.[3]

[3] Coate and Loury (1993) develop a more general model along these lines. They provide sufficient conditions for a patronizing equilibrium to exist.

Conclusions

The point of this exercise has been to illustrate, with the aid of formal economic reasoning, that the concerns expressed by some critics of quota-like anti-discrimination enforcement policies should be taken seriously. *My main result in these two simple models of worker-employer interaction is that, even when minority and majority groups have equal abilities on average, requiring equal representation of their members in high-level positions may distort incentives so as to produce an unintended consequence. Specifically, minorities may underinvest in the skills needed to perform adequately in such positions, relative to the investment rate of majority workers. That is, policies intended to assure equality of achievement may end up producing inequality of skills.*

The analysis suggests two general conclusions. First, the asymmetry of information between employers and government concerning the qualifications of workers who may or may not be subject to discrimination constitutes a serious obstacle to effective anti-discrimination policies. In both the taste discrimination and the statistical discrimination models, it is the inability of the enforcement agent to prescribe detailed procedural employment methodologies that forces reliance upon quota-like techniques. As is clear from the results of this chapter, these techniques can backfire.

A second implication is that gradualism is preferable to radical intervention when attempting to correct for group disparities thought to be the result of discrimination. When discrimination does occur, it discourages skill acquisition. A radical effort to enforce equality of representation can thus create bottlenecks, since there may actually be a shortage of qualified minority workers. Intervention that aims for increased, though less than fully proportional, minority representation allows time for the pool of qualified minority workers to expand in response to the improved opportunities, after which the enforcement goal can be made more ambitious. In this way, the chance of producing unintended negative consequences can be minimized.

This chapter is not meant to be an attack on the practice of preferential treatment for minority workers. Whatever the political and legal merits of such policies, I have shown that there are circumstances, involving either invidious discrimination or rational but self-fulfilling employer stereotypes, when the use of quota-like policies can have desirable results. However, this is not necessarily the case. It is important therefore that we try to understand, in the many concrete circumstances in which preferences are now employed, just when the risks of generating negative unintended consequences of the sort I identify here are worth taking. Too often both advocates and critics are content to base their arguments entirely on first principles, without reference to the direct or indirect consequences of this contentious policy. Further

study will be required to identify practically significant cases that exemplify the effects uncovered here.

References

Arrow, Kenneth. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, ed. Orley Ashenfelter and Albert Rees, pp. 3–33. Princeton, NJ: Princeton University Press.

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

Coate, Stephen, and Glenn Loury. 1993. "Will Affirmative Action Policies Eliminate Negative Stereotypes?" *American Economic Review* 83, no. 5 (December): 1220–1240.

Epstein, Richard. 1992. *Forbidden Grounds: The Case against Employment Discrimination Laws*. Cambridge, MA: Harvard University Press.

Twelve

Meritocracy, Redistribution, and the Size of the Pie*

ROLAND BÉNABOU

THIS CHAPTER examines how ambiguous notions such as "meritocracy," "equality of opportunity," and "equality of outcomes" can be given a formal content and related to more standard economic concepts such as social mobility, income inequality, and efficiency. It then proceeds to examine how redistributive policies affect each of these criteria of social justice and economic performance. This is done using a dynamic, optimizing model of earnings determination that incorporates ability, effort, family background, educational bequests, and redistributive policies. Because of endogenous labor supply and missing credit markets, redistribution has both adverse and beneficial effects on investment and output.

Writers on distributive justice have put forward very different views of what an individual "deserves" or is "entitled to." At one end is Rawls (1971), who sees no moral justification for differences in welfare among individuals. Innate talent and socioeconomic background are equally arbitrary forms of luck, which in themselves merit no reward. Some inequality is necessary to provide incentives for people to produce, but it should be kept to the minimum level consistent with maximizing the welfare of the most disadvantaged individual. At the other end are libertarians such as Nozick (1974), who view individuals as entitled to the entire endowment with which they came into the world, comprising both their own qualities and whatever was inherited from parents or other altruistic donors.[1] Common perceptions of fairness fall between these two extremes, with the line often drawn between innate qualities of the individual, which are mostly seen as true merits, and inherited economic and social advantages, which are not. For instance, Loury (1981) states that "it is widely held that differences in ability provide

* Prepared for the conference on "Meritocracy and Inequality" organized by the MacArthur Foundation at the University of Wisconsin–Madison. I am grateful to J. P. Benoît, Jason Cummins, Jordi Gali, and especially Efe Ok for useful suggestions and references. Financial support from the National Science Foundation (SBR-9601319) and the C. V. Starr Center is gratefully acknowledged.

[1] With the proviso that the capital thus transmitted should not have been acquired unjustly in the past, through expropriation, exploitation, or the like. But while Nozick briefly concedes that a principle of "just redress" is necessary, he remains remarkably silent on what it should be.