This article presents a model involving employers and two classes of workers, alike except for labels. Employers choose whom to hire and workers choose whether to invest in training. At one equilibrium, employers discriminate, which, the authors show, is Pareto inferior to another equilibrium where no discrimination occurs. On the basis of this observation, an argument for affirmative action is advanced.

An Economic Argument for Affirmative Action

DEAN P. FOSTER
RAKESH V. VOHRA
University of Chicago

In an unregulated, competitive market, those who discriminate for irrational and noneconomic reasons will be driven out of business. Thus any attempt to regulate a free market will hinder its ability to eliminate discrimination. Discriminatory practices that persist do so because they are efficient. Attempts to forbid efficient discrimination will reduce the total wealth of the community. This in a nutshell is the economic argument against preferential treatment, called here the market argument (given in detail in Posner 1981).

Although preferential treatment may be a bad thing, the market argument does not prove it, as shown with a simple and uncomplicated model. If the market argument fails here, then its validity in the larger world is suspect. As a corollary, we obtain an economic argument for preferential treatment. Because this is the climax of our tale, it is described after we have introduced the model.

Ours is not the first article concerned with providing an economic model of discrimination. It is a subject that has attracted the attention of many: Arrow (1972), Becker (1971), Marshall (1974), Milgrom (1987), and Phelps (1972), to name a few. A common feature of these models is a built-in asymmetry. For example, in Becker (1971), employers are assumed to have

RATIONALITY AND SOCIETY, Vol. 4 No. 2, April 1992 176-188

a taste for discrimination. In Phelps (1972), discrimination is the product of employers' skewed beliefs. In Milgrom (1987), certain employees are not as adept at "tooting their horns" as others. In short, the murderer is revealed first, while the details of the crime fill up the pages that remain. Furthermore, it is hard to imagine that the particular tastes or proclivities that lead to discrimination in some of these models would survive the rigors of a competitive market in the long run. The distinction of our model is that it contains no asymmetries of taste, ability, or proclivities.

THE MODEL

The model is a game between employers and potential employees. Potential employees are divided into two equal-sized groups: Alphas and Omegas. They are identical in all respects except that the Alphas have an α on their foreheads and the Omegas and Ω, which are clearly visible. Each potential employee can choose to invest, at a cost to his- or herself, in improving oneself. Only the potential employee knows for sure whether he or she has made this investment. A potential employee who makes such an investment will be called an I-worker. We do not allow the potential employee to choose the level of investment. It is a 0-1 decision. If they choose to invest, they improve themselves by a fixed amount (that is the same for everyone), otherwise, they do not. The qualitative conclusions of the model are not affected by this assumption.

The payoff to the potential employee under different conditions is as follows:

	If Hired	If Not Hired
Invest	1	0
Do Not Invest	a	b

Here, $a > 1 > b > 0$. The potential employee's objective is to select the strategy (invest or do not) that maximizes his or her payoff. Other things being equal, no employer is preferred to another. We emphasize that the payoffs to Alphas and Omegas are identical for identical circumstances. This ensures that any discrimination in the model is not the product of differences between Alphas and Omegas.

The employers in our model are identical and, other things being equal, have no preference for Alphas over Omegas or vice versa. However, each would prefer to hire an I-worker over a non-I worker because there is a significant positive correlation between self-investment and output. For the sake of simplicity, we assume there is no limit to the number of workers that may be hired.

If an employer hires an I-worker, his or her payoff is $h > 0$ and $h^* < 0$ if they hire a non-I worker. To help in identifying I-workers, the employers use a common test. The test is not perfectly reliable—no test can be. However, it is unbiased (this preserves the symmetry of the model). By this we mean that the scores obtained on the test depend only on whether one is an I-worker or not and not on whether one is an Alpha or Omega. It is this unbiasedness condition that distinguishes our model from a similar one suggested by Lundberg and Startz (1983), who assumed that the test is better at identifying I-workers among the Alphas than among the Omegas. It is difficult to see how such a test would not be replaced in the long run by a test that was more reliable for both Alphas and Omegas; employers would be better off using that test.

Each employer attempts to select a threshold/passing score on the test so as to screen out as many non-I's as possible and screen in as many I-workers as possible. Assume the test is scaled so that the range of potential test scores is in (0, 1). A threshold of 1 means that an employer does not wish to consider anyone for employment. Employers are at liberty to set different thresholds for different groups. The only condition that we impose on the test is that it must be *informative*. By this we mean there is a threshold at which the expected payoff to a potential employee from investing is larger than the expected payoff from not investing. Without this minimal condition, there would never be an incentive to take the test.

All transaction costs involved with administering the test, applying for a job, and so on are zero. All actors in our little drama are risk neutral.

Our model is different from a similar one proposed by Arrow (1972). In the Arrow model, employers discriminate on the basis of wages; different races can receive different wages for the same work. In the current climate, it would be difficult to find such a practice being widespread. More reasonably, one would expect that any discrimination that does occur would occur in hiring (equivalently, promotion) decisions. This is what our model attempts to capture.

While this article was under review, an as yet unpublished manuscript by Coate and Loury (1991) came to our attention. The model presented there is essentially the same as ours. The main differences are that potential employees can choose the level of self-investment and the cost of such an investment is stochastic. Nevertheless, all qualitative features exhibited by the Coate and Loury model are exhibited by our simpler model.

THE ANALYSIS

We start by considering the problem faced by the employers: the selection of an optimal test threshold. Assume, to begin with, a homogeneous population of potential employees. Let p be the probability that a randomly selected potential employee is an I-worker. Let f(x) = Pr(a worker scores x| worker is an I) and f*(x) = Pr(a worker scores x| worker is non-I). Let

$$F(x) = \int_0^x f(t)dt$$

and

$$F^*(x) = \int_0^x f^*(t)dt.$$

Hence the probability that a potential employee drawn at random is an I-worker and scores less than x is pF(x). Without loss of generality, we assume that f(x)/f*(x) is monotone (i.e., the higher the score on the test, the higher the chance that one is an I-worker) and continuous.

The expected payoff to an employer from selecting a threshold of x is

$$hp(1 - F(x)) + h^*(1 - p)(1 - F^*(x)).$$

The optimal threshold, t, is the value of x at which the expression above attains its maximum. To find this, we differentiate and set equal to zero. Thus t must satisfy

$$-hpf(t) - h'(1 - p)f^*(t) = 0. \tag{1}$$

A solution exists because f(x)/f*(x) is monotone and continuous. The monotonicity of f(x)/f*(x) also ensures that t maximizes the employer's payoff.

Turn now to a potential employee, whose expected payoff from investing is

$$\int_t^1 f(x)dx \tag{2}$$

and from not investing is

$$a\int_t^1 f^*(x)dx + b(1 - \int_t^1 f^*(x)dx). \tag{3}$$

OBSERVATION 1

Consider what happens when $t = 1$ (i.e., no one is considered for employment). Then, the expected payoff from not investing is $b > 0$ (= the expected payoff from investing). Hence no potential employee invests, and so no employer has an incentive to change the value of t. We have a Nash equilibrium where no one is hired and no one invests.

OBSERVATION 2

Let τ be selected so that if $t = \tau$, Equation 2 will be equal to Equation 3 (i.e., the expected payoff to a potential employee from investing is equal to the payoff from not investing). Such a τ exists because the test is informative (τ need not be unique). Let π be such that $-h\pi f(\tau) - h^*(1 - \pi)f^*(\tau) = 0$. In other words, π is selected so that τ will be the optimal threshold for each employer. Notice that (τ, π) form a Nash equilibrium. To see why, observe that as the expected payoff from investing and not investing is the same, no potential employee has an incentive to deviate. Keeping π fixed, τ is the optimal threshold for each employer. So, they have no incentive to change it, either. It is interesting to note that this Nash equilibrium is Pareto-superior to the one of Observation 1.

A graphical illustration of these observations is provided. In Figure 1, we have a graph of the probability of passing the test, $1 - F(x)$ (vertical axis) at a given threshold (horizontal axis) for both I's and non-I's. We have normalized the threshold so that the curve for a non-I is a straight line with slope −1. From these two curves we can construct the curves of Figure 2. Here, we plot expected payoff to a potential employee versus threshold for both I's and non-I's. The possible values of τ are labeled t' and t''. Figure 3 shows a plot of $hp(1 - F(x))$ and $h^*(1 - p)(1 - F^*(x))$ against threshold t. The graph of expected payoff to a an employer against threshold is found by adding the two curves of Figure 3 together. The results are shown in Figures 4a and 4b.

Notice what happens when all employers choose $t = 1$ as a threshold. From Figure 2 we see that the expected cost of not investing exceeds that of investing. Hence no potential employee will invest (i.e., $p = 0$). From Figure 3 with $p = 0$ (so the curve in the nonnegative orthant does not exist) we see that if the employer sets a threshold $t < 1$, he or she reduces his or her payoff. Clearly, any party that deviates unilaterally from this arrangement will be worse off.

Suppose now that all employers choose t'' as their threshold. From Figure 2 we see that the expected payoff from investing is equal to that from not investing. Choose a $p'' \varepsilon (0, 1)$ so that $-hp''f(t'') - h'(1 - p'')f^*(t'') = 0$. If any

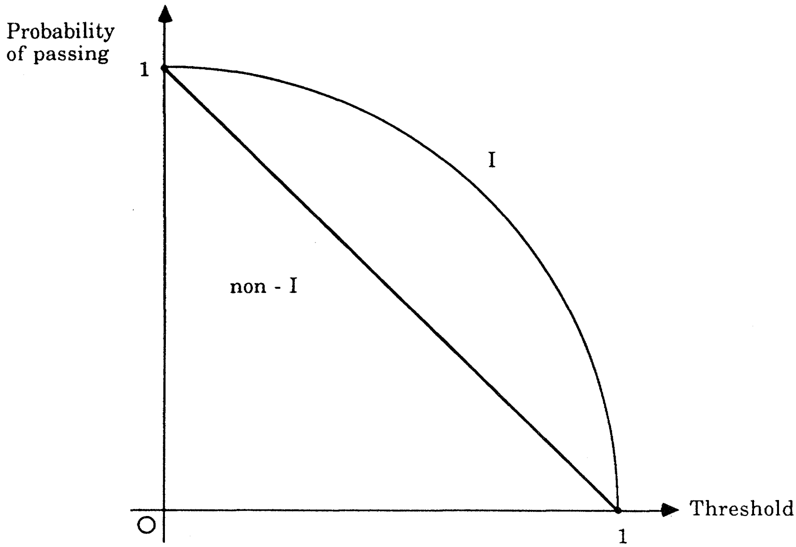

Figure 1

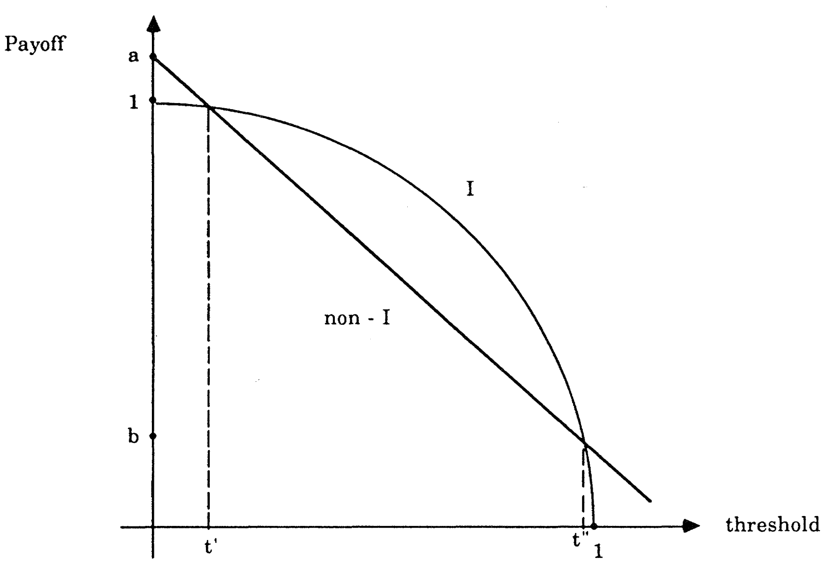

Figure 2

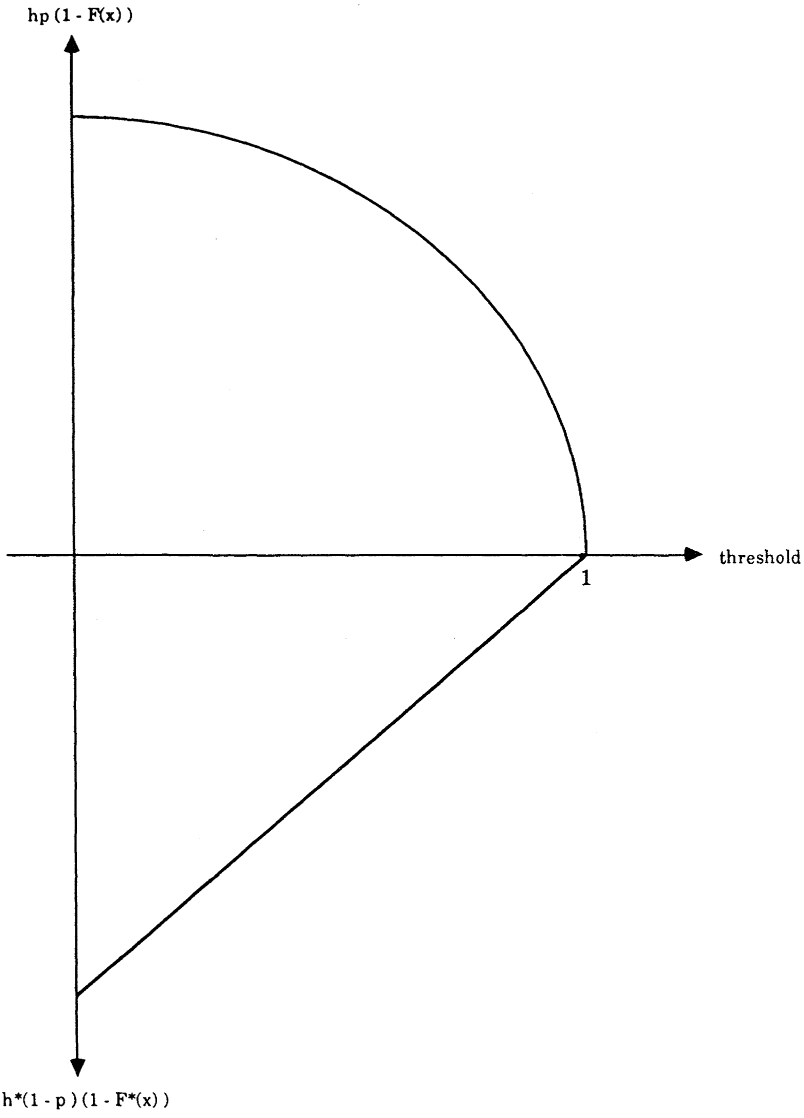

Figure 3

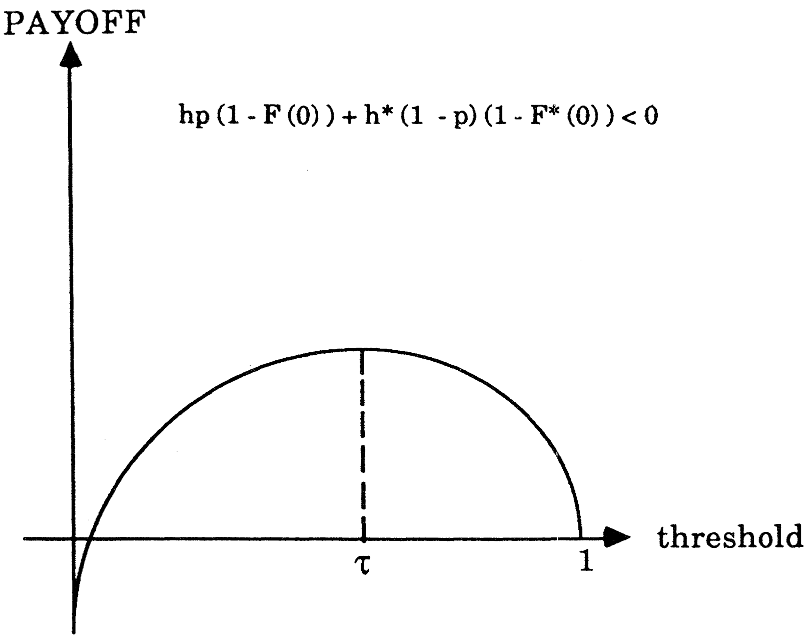

Figure 4a

employer chooses a t different from t″, they will be worse off because for the selected value of p″, t″ is the choice of threshold that maximizes an employers payoff (see Figure 4). No potential employee has an incentive to deviate because the expected payoff from either of their strategies is identical. Thus (p″, t″) is a Nash equilibrium. A similar analysis applies to t′.

It is important to note that of the Nash equilibria centered at t′ and t″, only the one at t′ is stable. To see why, consider what happens if the employers choose a threshold slightly higher than t′. Then, it is in each potential employee's interest to invest. The employers can now reduce their threshold and increase their payoffs because they would be passing more I-workers. If the employers select a threshold slightly lower than t′, the potential employees will choose not to invest. This makes the employers worse off, so they increase their threshold. To see why the equilibrium at t″ is not stable, consider what happens if the employers choose a threshold slightly lower than t″. Then, all potential employees will invest, and so the employers would be better off picking t′ as their threshold.

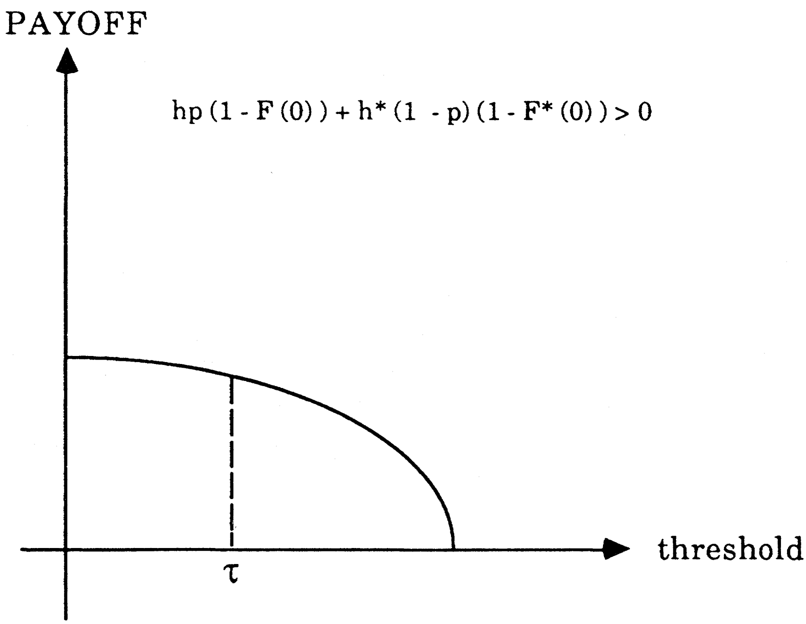

Figure 4b

Now, return to the situation where the set of potential employees consists of both Alphas and Omegas. The following set of strategy choices constitute an equilibrium by Observations 1 and 2:

1. Choose p and t with respect to the Alphas as in Observation 2.
2. Set t = 1 for the Omegas, and all Omegas choose not to invest.

Call this the discriminatory equilibrium (DE).

To see what happens if an employer chooses a pair of thresholds (one for Alphas, the other for Omegas) other than those that form an equilibrium, see Figure 5a. A pair of thresholds is represented by a point in the plane. Suppose the pair selected is in the upper left-hand cell. Then, the employer can improve his or her payoff by changing the thresholds in the direction of the arrow lying in that cell – in this case, to (t′, 1) (i.e., the DE). The cells shown are drawn to be of equal size for aesthetic reasons only.

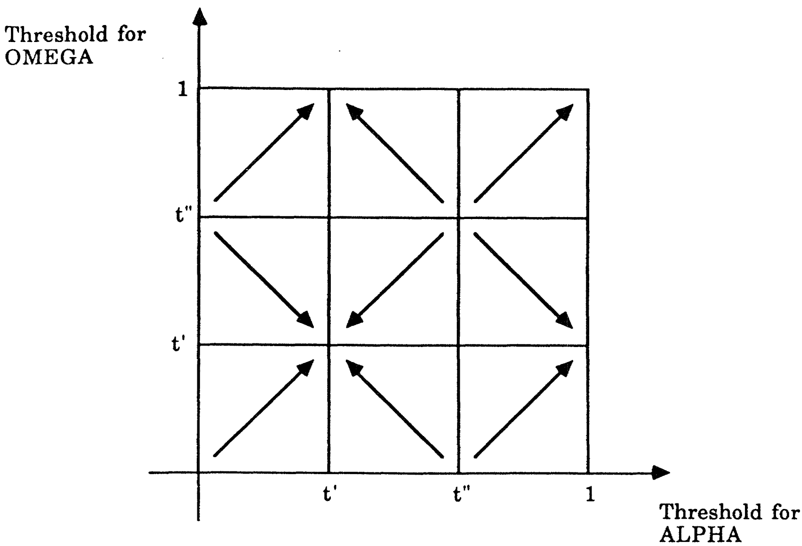

Figure 5a

DISCUSSION

At first glance, existence of the DE does not contradict the market argument. The employers discriminate against the Omegas, but it is rational for them to do so as none of the Omegas choose to invest. Notice, however, that the DE is Pareto-inferior to the equilibrium obtained by applying Observation 2 to both Alphas and Omegas (the Lundberg and Startz [1983] model only guarantees an increase in total welfare in this instance). Thus the discrimination, while rational for the employers, is not efficient for the society as a whole.

The weakness in the market argument is that it looks at the matter through the eyes of the employer only (see Conway and Roberts [1983] for a similar notion in different language). It ignores the incentive effects that an employer's action has on a potential employee. Employers do not consider Omegas because they do not invest; Omegas do not invest because they will never be considered for employment—catch-22. *The inefficient discrimination occurs because the incentives for Alphas and Omegas are not the same.*

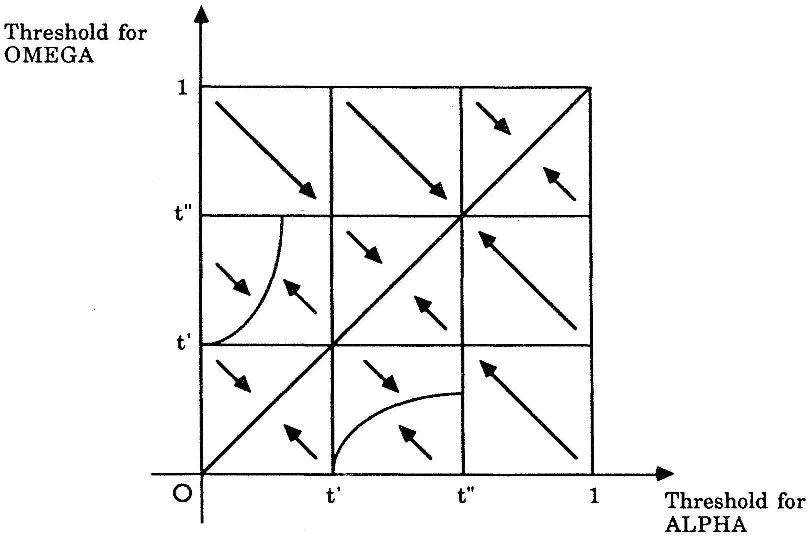

Figure 5b

How does one get out of the DE? Improving the reliability of the test does not do this; it just increases the number of Alphas who would be employed. The trick is to structure the incentives for the Omegas so that it is rational for them to invest. Within the context of the model, there is an easy way to do this: require employers to use the same threshold for all potential employees.

To study the effect of the imposition of an equal threshold policy when not in the DE, see Figure 5b. Imagine that a fine, proportional to the absolute difference in the number of Alphas and Omegas employed, is imposed on all employers. If the employer has selected a pair of thresholds in the upper left-hand cell, he or she is better off moving in the direction of the arrow in that cell. In other words, the employer will move to equalize the thresholds for Alphas and Omegas.

An interesting thing happens if an employer has selected a pair in the middle cell of the third row. If the corresponding point is below the curve in that cell, the employer will be better off equalizing the thresholds. If the point is above the curve, the employer would move in the opposite direction. For a particular choice of a, b, h, h*, π, F, F*, and fine, there is an equilibrium where (a) the threshold for Omegas is lower than the threshold for Alphas, (b) all Alphas invest and none of the Omegas do, and (c) the same proportion of Alphas and Omegas are hired. In plain language, Alphas and Omegas are hired in equal numbers, but the Omegas meet a lower standard. Coate and Loury (1991) called such an equilibrium a patronizing one (PE). The hypothetical fine used to generate Figure 5b is a Lagrange multiplier in the Coate-Loury model.

The important point to note is that the DE marked by the point (t′, 1) and the PE (middle cell, last row) are on opposite sides of the 45° line. The 45° line represents a situation where both groups are treated identically. Once there, there are no forces to push one off the 45° line into a PE.

IMPLICATIONS

Quotas, affirmative action, and other forms of preferential treatment have been the subject of much controversy, particularly in light of the current debate over the new civil rights bill (see, e.g., *New York Times* of 29 May, 1991, and *Time* magazine of 27 May, 1991). Although many arguments in defense of preferential treatment have been offered, they have been unconvincing (see Posner 1981 for a strong and spirited criticism of them). Preferential treatment does not appear to treat people equally. This is the hurdle that any convincing argument in its defense must clear.

The preceding analysis suggests such an argument. Specifically, the purpose of preferential treatment is to ensure that all groups of potential employees face *identical incentives*. We illustrate the use of this argument to provide a justification for a form of preferential treatment called race norming. In race norming, a candidate is ranked on the basis of his or her test score and the scores of other test takers of the same race. An employer then uses the same rank cutoff for all potential employees. Thus, for an Omega to be ranked in the 98th percentile, say, it is sufficient that he or she score higher than 98% of the other Omegas taking the test.

Recall, that if one is in the DE, a way out is to insist that all employers use the same threshold for Alphas and Omegas. This is not enough. Suppose that the employers could switch from the hypothetical unbiased test of the model to one that could distinguish between Alphas and Omegas (i.e., Omegas consistently score lower than Alphas). It would be in their interests to do so. They could set the same threshold for Alphas and Omegas and be confident that Omegas would score below the threshold. In other words, they maintain the DE. Of course, one could avoid this by prohibiting the use of biased tests. The difficulty with this is a practical one. We know of no test that is culturally unbiased and, more important, perceived as such. In the absence of a culturally unbiased test, what is to be done?

In this context, fairness requires that the test predict self-investment for Alphas and Omegas with the same reliability. In other words, the incentives for Alphas and Omegas to invest in themselves must be the same. This can be achieved by ensuring that the false positive rate of the test for Alphas and Omegas be the same. A simple and practical way of achieving this is through race norming.

CONCLUSION

The model in this article has made two points:

1. Discrimination, while individually rational for the employer, may be socially inefficient.
2. The cause of this inefficient discrimination is the inequality in incentives for potential employees.

The two arguments provide an economic justification for intervention to correct this inefficiency. Such intervention will take the form of quotas, race norming of tests, or other kinds of preferential treatment. Also, such preferential treatment is fair, as it ensures that potential employees of all groups receive the same rewards for the same investment.

REFERENCES

Arrow, Kenneth J. 1972. Models of job discrimination. In *Racial discrimination in economic life*, edited by A. H. Pascal. Lexington, MA: Lexington Books.

Becker, Gary S. 1971. *The economics of discrimination*. Chicago: University of Chicago Press.

Coate, S., and G. Loury. 1991. Will Affirmative action policies eliminate negative stereotypes? Manuscript.

Conway, Delores, and Harry V. Roberts. 1983. Reverse regression, fairness, and employment discrimination. *Journal of Business and Economic Statistics* 1:75-85.

Lundberg, Shelly J., and Richard Startz. 1983. Private discrimination and social intervention in competitive labor markets. *American Economic Review* 73:340-47.

Marshall, Ray. 1974. The economics of racial discrimination. *Journal of Economic Literature* 12:847-71.

Milgrom, Paul. 1987. Job discrimination, market forces and the invisibility hypothesis. *Quarterly Journal of Economics* 102:435-76.

Phelps, E. S. 1972. The statistical theory of racism and sexism. *American Economic Review* 62:659-61.

Posner, Richard A. 1981. *The economics of justice*. Cambridge, MA: Harvard University Press.