The Statistical Theory of Racism and Sexism

By Edmund S. Phelps[*]

My recent book, *Inflation Policy and Unemployment Theory*, introduces what is called the statistical theory of racial (and sexual) discrimination in the labor market.[1] The theory fell naturally out of the non-Walrasian treatment there of the labor "market" as operating imperfectly because of the scarcity of information about the existence and characteristics of workers and jobs.

A paradigm for the theory is the traveller in a strange town faced with choosing between dinner at the hotel and dinner somewhere in the town. If he makes it a rule to dine outside the hotel without any prior investigation, he is said to be discriminating against the hotel. Though there will be instances where the hotel cuisine would have been preferable, the rule represents rational behavior—it maximizes expected utility— if the cost of acquiring evaluations of restaurants is sufficiently high and if the hotel restaurant is believed to be inferior at least half the time.

In the same way, the employer who seeks to maximize expected profit will discriminate against blacks or women if he believes them to be less qualified, reliable, long-term, etc. on the average than whites and men, respectively, and if the cost of gaining information about the individual applicants is excessive. Skin color or sex is taken as a proxy for relevant data not sampled. The a priori belief in the probable preferability of a white or a male over a black or female candidate who is not known to differ in other respects might stem from the employer's previous statistical experience with the two groups (members from the less favored groups might have been, and continue to be, hired at less favorable terms); or it might stem from prevailing sociological beliefs that blacks and women grow up disadvantaged due to racial hostility or at least prejudices toward them in the society (in which latter case the discrimination is self-perpetuating).

The theory is applicable to the class of "liberal" employers and workers who have no distaste for hiring and working alongside black or female workers. By contrast, the theory of discrimination originated by Gary Becker is based on the factor of racial taste. The pioneering work of Gunnar Myrdal et al. also appears to center on racial (and, in an appendix, sexual) antagonism.

Some indications of interest in the new theory, and the independent discovery of the same statistical theory by Kenneth Arrow, convince me that it is time for a formalization of the theory in terms of an exact statistical model. Though what follows is very simple, it may be useful to those who like exact models and it may stimulate others to develop the theory further.

An employer samples from a population of job applicants. The employer is able to measure the performance of each applicant in some kind of test, y_i, which, after suitable scaling, may be said to measure the applicant's promise or degree of qualification, q_i, plus an error term, μ_i.

$$(1) \qquad y_i = q_i + \mu_i$$

where μ is normally distributed with mean zero.

It is conceivable (and it sometimes occurs in practice) that the employer will have no other information about each applicant, including skin color.[2] In that special case, the employer may use q_i as a least-squares predictor of the applicant's y_i according to the regression-type relation:

---

[2] The Fair Employment Practices Law forbids employers from asking for information on race in written applications. The Boston Symphony Orchestra auditions candidates from behind an opaque screen.

$$q_i' = a_1 y_i' + u_i'$$

(2)
$$0 < a_1 = \frac{\text{var } q_i'}{\text{var } q_i' + \text{var } \mu_i} < 1, \; Eu_i = 0$$

where q_i' and y_i' are deviations from their respective population means.[3]

Suppose instead that skin color is observed along with the test datum, and suppose that the employer postulates a model of job qualification

(3)
$$q_i = \alpha + x_i + \eta_i$$

in which

(3a) $x_i = (-\beta + \epsilon_i)c_i, \quad \beta > 0,$

where $c_i = 1$ if the applicant is black and zero otherwise. Here x_i is the contribution of social factors, and these are believed to be race-related according to (3a). The random variables ϵ_i and η_i are normally and independently distributed with mean zero. Letting $\lambda_i = \eta_i + c_i\epsilon_i$ and $z_i = -\beta c_i$, we may write

(4)
$$q_i = \alpha + z_i + \lambda_i$$
$$y_i = q_i + \mu_i = \alpha + z_i + \lambda_i + \mu_i$$

Then the test datum can be used in relation to the race (sex) factor to predict the degree of qualification net of the race factor, the latter being separately calculable:

(5) $q_i' - z_i' = a_1 \cdot (y_i' - z_i') + u_i$

$$0 < a_1 = \frac{\text{var } \lambda}{\text{var } \lambda + \text{var } \mu} < 1$$

or, equivalently

(5') $q_i' = \dfrac{\text{var } \lambda_i}{\text{var } \lambda_i + \text{var } \mu_i} \cdot y_i'$

$$+ \frac{\text{var } \mu_i}{\text{var } \lambda_i + \text{var } \mu_i} \cdot z_i' + u_i$$

[3] In (2), a_1 is the probability limit, as $N \to \infty$, of the regression coefficient

$$\hat{a}_1 = \frac{1}{N}\sum_{i=1}^{N} y_i' q_i' \Big/ \frac{1}{N}\sum_{i=1}^{N} (y_i')^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} (q_i' + \mu_i)q_i' \Big/ \frac{1}{N}\sum_{i=1}^{N} (q_i' + \mu_i)^2$$

The weights applied to the test information and the skin color information are inversely related to the variances of the respective disturbance terms corresponding to them.[4]

CASE 1. If growing up black is believed by the employer to be socially disadvantageous, so that $z_i' < 0$ for black applicants, then one might expect to find a lower prediction of q_i for blacks than whites having equal test scores. This is generally true, however, only in the special case where $\epsilon_i \equiv 0$ for all i, i.e., for all blacks as well as whites. This means that there is no differential variability in promise as between blacks and whites. Then var $\lambda_i =$ var η_i and hence the coefficients in (5') are independent of c_i. Therefore the prediction curve relating q_i to y_i for blacks lies parallel and below that for whites, as illustrated in Figure 1.
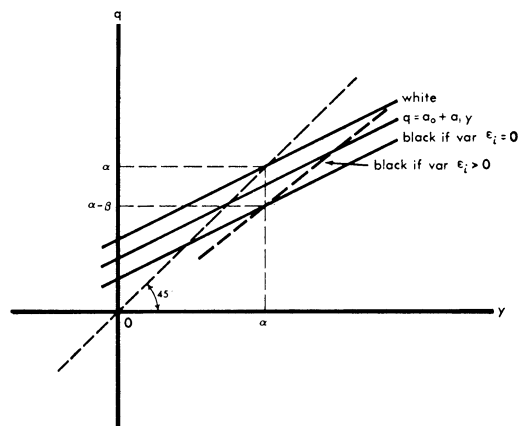

FIGURE 1. PREDICTION OF QUALIFICATION BY RACE AND TEST SCORE

CASE 2. In general the variance of λ depends upon skin color. The formulation in (3) ascribes to blacks the larger postulated variance, as reflected in (6):

(6)   var $\lambda_i =$ var $\eta_i + c_i^2$ var ϵ_i

[4] My attention has been called by the referee to the derivation of a generalization of equation (5'), from which can be deduced all my cases, in the extended footnote on page 325 in Thomas Wonnacott and Ronald Wonnacott.

It follows that the coefficient of the test score in the least-squares prediction of qualification is *greater* for blacks than for whites. (In the limit, as var $\epsilon_i \to \infty$, the coefficient of y_i—the slope of the prediction curve for blacks—approaches one.) For any positive var ϵ_i it is a consequence of the race-related difference in coefficients that at some high test score and higher ones the black applicant is predicted by the employer to excel over any white applicant with the same or lower score. The employer credits an equally good test score by the white applicant as a *less credible* indication in view of the prior notions of the comparatively narrow range of white promise. Note that one can reverse these implications by replacing the dummy variable in (3) with $(1 - c_i)$ instead.

A FURTHER CASE. It is straightforward to make the disturbance term in (1) conditional on race in the way that λ was made conditional on skin color:

$$(7) \qquad \mu_i = \xi_i + c_i \rho_i$$

Then whites' test scores are regarded by the employer as *more reliable* than the scores of blacks—that is, they measure promise with less error. In that case the greater reliability of whites' test scores might overcome any tendency for them to have less credibility, so that the white prediction curve would be the steeper curve. Then there is a range of low test scores in which whites are predicted to be less qualified than equally high scoring blacks.

A final word. A sensitive person, I have been warned, might read this paper as expressing an impression on the part of the author that most or all discrimination is the result of beliefs that blacks and women deliver on the average an inferior performance. Actually, I do not know (nor claim to know) whether in fact most discrimination is of the statistical kind studied here. But what if it were? Discrimination is no less damaging to its victims for being statistical. And it is no less important for social policy to counter.

REFERENCES

K. J. Arrow, "Some Models of Racial Discrimination in the Labor Market," RAND Corporation research memorandum RM-6253-RC, multilith, Santa Monica, Feb. 1971.

G. S. Becker, *The Economics of Discrimination,* Chicago 1959.

G. Myrdal, *An American Dilemma,* New York 1944.

E. S. Phelps, *Inflation Policy and Unemployment Theory,* New York 1972.

T. H. Wonnacott and R. J. Wonnacott, *Introductory Statistics,* New York 1969.